

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA
Corso di Laurea Magistrale in Ingegneria Informatica

Anomaly Prediction in Production Supercomputer with Convolution and Semi-supervised autoencoder

Relatore:
Prof.ssa
Michela Milano

Candidato:
Massimo Schembri

Correlatore:
Dott.
Andrea Borghesi

IV Sessione
Anno Accademico
2019/2020

Indice

Introduzione	5
1 Infrastruttura	7
1.1 MARCONI	7
1.2 Examon	7
1.2.1 Plugins	8
1.2.2 Estrazione e manipolazione dei dati	9
2 Anomaly Detection	11
2.1 Tecniche	11
2.1.1 Autoencoders	12
2.2 Related works	13
3 Progetto	15
3.1 Autoencoder semisupervised	15
3.1.1 Costruzione di train e test set	16
3.2 Analisi dei grafici	18
3.2.1 Analisi di Gennaio	18
3.2.2 Analisi di Maggio	55
3.3 Modello ensemble e detection	98
3.3.1 Detection	98
3.3.2 Analisi modello ensemble v1	99
3.3.3 Analisi random false positives	128
3.3.4 Analisi modello ensemble v2 - Mean	130
3.3.5 Analisi modello ensemble v2 - Window Mean	139
3.3.6 Analisi modello ensemble v2 - WindowMean&sum	153
3.4 Confronto	176
4 Conclusioni	179
Bibliografia	181

Introduzione

HPC (High Performance Computing) è un acronimo non molto comune, almeno non quanto per esempio Big Data o IoT, ma non per questo ha un ruolo secondario all'interno del panorama scientifico.

Con HPC s'intende l'abilità di poter processare dati e poter effettuare operazioni complesse ad alte velocità. In particolare la velocità di questi sistemi viene misurata in FLOPS (Floating point operations per second), quindi le operazioni in virgola mobile che possono essere effettuati ogni secondo.

La soluzione HPC più conosciuta è il caso del supercomputer, un'infrastruttura che contiene migliaia di nodi interconnessi tra loro che lavorano assieme per portare a termine dei job [9].

I job possono essere sottomessi da vari utenti poiché questo tipo di infrastruttura può essere affittata da varie aziende per sfruttare questa potenza di calcolo per i propri scopi. Se un'azienda fa di quest'infrastruttura il proprio business, o è una parte fondamentale del proprio processo produttivo, è importante che l'infrastruttura sia disponibile il più possibile e che eventuali problemi siano risolti in tempi brevi o addirittura prevenuti.

Il processo che dà luogo a questa risoluzione dei problemi prima che accadano effettivamente viene chiamato "predictive maintenance" il quale obiettivo è quello di minimizzare rotture inaspettate dei componenti e sfruttare al meglio i componenti [10].

L'obiettivo di questa tesi è quello di analizzare e realizzare di un sistema di anomaly detection in grado non solo di riuscire a identificare con successo le anomalie di Marconi quando sono in corso, ma riuscire a prevederne l'avvenimento nei limiti del possibile. In modo tale da dare luogo alla "predictive maintenance".

Capitolo 1

Infrastruttura

In questo capitolo spiegherò il contesto in cui ho operato durante l'attività di tesi, il sistema HPC e l'elaborazione delle metriche di performance riguardante il sistema MARCONI e il sistema Examon che è stato fondamentale per lo sviluppo di quest'attività.

1.1 MARCONI

Il sistema HPC sul quale è stata effettuata quest'attività di tesi è appunto MARCONI di CINECA.

Il sistema MARCONI è nato nel 2016 e durante il proprio periodo operativo ha subito vari upgrade fino a raggiungere la diciannovesima posizione nella classifica dei sistemi HPC più potenti al mondo con 20 PFLOPS di capacità di calcolo e 17 PB di storage.

L'architettura di MARCONI è proprio come quella descritta nella sezione d'introduzione, ovvero quella a supercomputer. MARCONI è quindi formato da migliaia di nodi interconnessi tra loro.

1.2 Examon

Examon è un framework altamente scalabile che permette di monitorare le performance di un sistema HPC.

La struttura di Examon è rappresentata in figura 1.1, nella quale si possono osservare più livelli, a partire dal singolo nodo di MARCONI, fino ad arrivare alle applicazioni generiche che possono manipolare e usufruire dei dati forniti dal sistema Examon.

La comunicazione da nodo HPC al database è realizzata tramite protocollo MQTT che è ottimo per scenari nei quali è richiesta un uso minimo

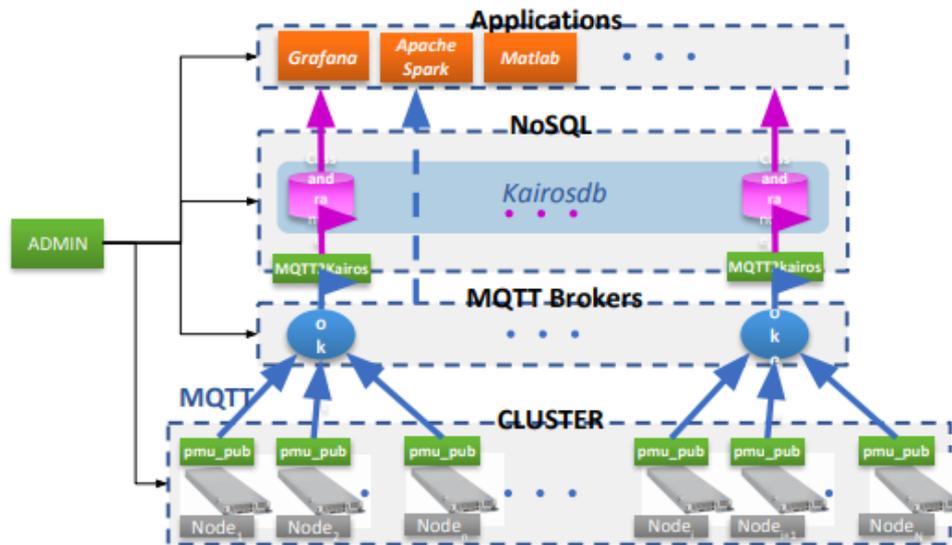


Figura 1.1: Struttura di examon, Fonte: slide

di risorse. L'utilizzo di questo protocollo permette di utilizzare un modello pub/sub e una comunicazione asincrona dei messaggi con un basso overhead.

I dati raccolti dai nodi HPC vengono infine memorizzati su un database NOSQL, KairosDB in questo caso, che sarà poi accessibile all'esterno per poter reperire i dati in questione.

1.2.1 Plugins

L'architettura di Examon permette di delegare la raccolta dei dati ad altri componenti software denominati plugin. Questa modularità del sistema permette di poter aggiungere oppure rimuovere le metriche da monitorare e di cui conservare i dati nel database. Inoltre nel database è anche presente l'origine, in termini di plugin, di questi dati in questo modo è possibile recuperare i dati relativi a uno specifico plugin [4].

I plugin e i relativi dati utilizzati in questa tesi sono i seguenti:

- **Ganglia:** Questo plugin si occupa di raccogliere informazioni riguardante la CPU e il carico del nodo. Per esempio frequenza della CPU e RAM libera.

- **Confluent:** Questo plugin invece si occupa di tutto ciò che riguarda gli aspetti energetici del sistema, quindi temperatura dei componenti, utilizzo di rete e disco fisso.
- **Nagios:** Questo plugin ha il compito di raccogliere dati sullo stato del nodo. In particolare i dati riguardanti questo plugin vengono inseriti manualmente allegati a una descrizione testuale, quando è presente un'anomalia viene inserito un numero da 0 a 3 riguardante il livello di urgenza. Nel nostro caso considereremo anomalia solamente i valori sopra il 2 e che nel testo contengano "DOWN+DRAIN" consigliato dopo un'analisi con i system admin.

I dati salvati dei plugin possono essere riassunti in breve in tre valori: timestamp, metrica, valore. Occorre precisare anche che ogni metrica può avere una frequenza di campionamento differente dalle altre.

1.2.2 Estrazione e manipolazione dei dati

L'estrazione dei dati è stata effettuata tramite script Python ed è stata possibile grazie alla libreria ExamonQL di Francesco Beneventi.

Con la quale è possibile interrogare il database di Examon tramite un linguaggio molto simile a SQL, rendendo semplice recuperare i dati in quanto vengono restituiti come Pandas DataFrame, una struttura molto comune e utile in questo ambito.

Una fase di preprocessing risulta necessaria una volta ottenuti i dati dal database poiché ogni riga, come detto precedentemente, è formata da timestamp, metrica, valore. Il primo problema è appunto la diversa frequenza di campionamento di ogni metrica e il fatto che il timestamp è preciso al millisecondo e ovviamente non si hanno sempre timestamp uguali fra loro.

L'obiettivo di questa fase di preprocessing è quindi quella di creare un singolo record che raccolga tutte le metriche necessarie, osservando le varie frequenze di campionamento si è scelto di riassumere in un singolo record i dati presenti in una finestra di cinque minuti in quanto coincide con la frequenza di campionamento dello stato del nodo fornito dal plugin NAGIOS che rappresenterà le label del dataset.

Nel nostro caso abbiamo deciso di riassumere questa finestra di 5 minuti tramite media e varianza dei valori raccolti da Examon, questa operazione consente anche di ridurre la dimensione del dataset finale rispetto ai dati originariamente estratti da Examon.

Per l'attività che è stata svolta questo passo risulta fondamentale e necessario per creare un dataset su cui effettuare analisi e training di reti neurali.

Capitolo 2

Anomaly Detection

L'attività di anomaly detection è nata per rilevare istanze all'interno di un'insieme che si discostano dalle altre, queste istanze vengono appunto chiamate anomalie. La rilevazione delle anomalie viene fatta in maniera del tutto data-driven, ovvero si cerca di modellare come dovrebbe essere un'istanza a regime e si osserva quanto una nuova istanza è distante da questo modello.

Questo tipo di attività può essere svolta in molti ambiti, per esempio: rilevazioni di frodi fiscali, attività di monitoraggio di qualsiasi genere. Ciò lascia intendere la rilevanza dell'argomento ai giorni nostri [5].

L'attività di anomaly detection si basa su due ipotesi:

- La percentuale di istanze anomale all'interno dell'insieme è molto bassa, anche perché se così non fosse non si parlerebbe più di istanze che differiscono dalle altre.
- A livello di features presentano pattern che consentono di classificare queste istanze come anomale.

2.1 Tecniche

Sono presenti differenti tecniche per effettuare questo tipo di attività che dipendono anche dal tipo di istanze che si hanno a disposizione, ovvero se sono presenti etichette, inserite a posteriori da un esperto, sulle istanze che sono anomale. Le tecniche le possiamo racchiudere in tre macro aree: supervised learning, semi-supervised learning, unsupervised learning.

Il supervised learning può essere utilizzato come tecnica solamente se il nostro set d'istanze è stato etichettato. Questo approccio presenta una criticità nell'ambito dell'anomaly detection, poiché come elencato nelle ipotesi, le anomalie sono presenti in modo molto più raro rispetto alle istanze a regime.

Questo porta ad avere un modello detto "biased" in gergo, ovvero che è più inclinato verso la classe che è più numerosa all'interno del dataset e questo potrebbe essere la causa di un modello con performance basse in questo campo. Può inoltre succedere che la percentuale di record anomali riguardi anomalie di natura differente, in questo caso la rete potrebbe avere scarse performance nel rilevare i vari tipi di anomalia.

Unsupervised Learning con questa tecnica si cerca di estrarre i pattern presenti nei dati al fine di raggruppare le istanze in gruppi nei quali queste istanze hanno tratti comuni. In gergo questi gruppi vengono detti clusters. Questa tecnica non richiede la presenza di label all'interno del set d'istanze.

Semisupervised Learning è una tecnica che può essere utilizzata se si hanno a disposizione le label del set d'istanze. Solitamente questa tecnica viene applicata da un modello di rete neurale chiamato "autoencoder".

In questa tecnica viene fatto imparare all'autoencoder a riprodurre le istanze non anomale presenti nel set e le relazioni tra le varie features. In questo modo quando viene fornita un'istanza anomala, grazie all'ipotesi formulate precedentemente, dovrebbe mostrare un errore di ricostruzione su qualche feature, in questo modo è possibile individuare anomalie.

Nella sezione seguente espanderò meglio cosa s'intende per autoencoder.

2.1.1 Autoencoders

Un autoencoder è un tipo specifico di rete neurale con il compito di codificare l'input in una rappresentazione significativa con dimensionalità minore e riuscire a decodificarla in modo tale che l'output ricostruito sia il più simile possibile all'input ricevuto[6]. Con questo obbiettivo in mente, è logico che la funzione da minimizzare (funzione di loss) per effettuare il training di questa rete sia data dall'errore di ricostruzione, ovvero la differenza tra output e input, solitamente al quadrato.

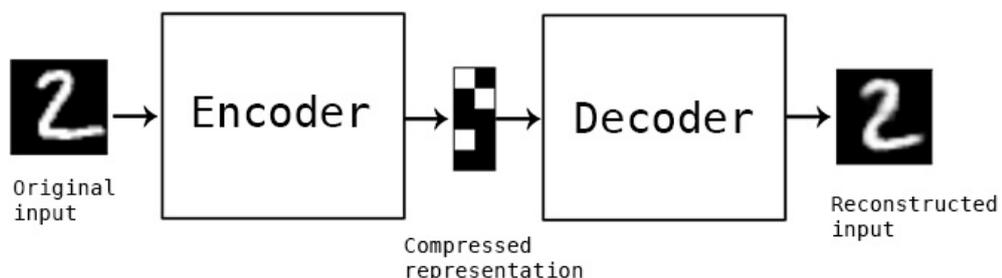


Figura 2.1: Struttura autoencoder [7]

Come si può vedere nella figura 2.1 e come detto precedentemente, un autoencoder può effettuare una riduzione di dimensionalità del dataset. Può essere utilizzato al posto di una tecnica come PCA che effettua un mapping di tipo lineare, mentre un autoencoder effettua un mapping di tipo non lineare quindi con un potenziale espressivo molto più alto.

Questo è un singolo caso d'uso della struttura ad autoencoder, può essere utilizzato per molti altri scopi ma in questa tesi è stato utilizzato essenzialmente nelle modalità esposte.

2.2 Related works

Un esempio di sistema di anomaly detection che utilizza un approccio unsupervised è il lavoro di Zhang et al. [1], nel quale viene appunto utilizzato un autoencoder in modo unsupervised e matrici che catturano le relazioni tra le varie features e poter effettuare una diagnosi del sistema.

Per quanto riguarda l'approccio supervised, anche se è accoppiato con altri metodi, è utilizzato nel lavoro di Wang et al. [2]. Nel quale un sistema di deeplearning, ancora una volta un autoencoder, vengono affiancate due tecniche PCA (unsupervised) e SVM(supervised). PCA per effettuare un preprocessing delle feature estratte dall'autoencoder e SVM come metodo di supervised learning che può fornire performance migliori di un layer di softmax in questo contesto.

Un aspetto molto importante è anche l'interpretabilità dei risultati, proprietà che spesso manca quando si utilizzano tecniche come deep learning. Nel lavoro di Baseman et al.[3] viene sfruttata l'espressività di Random forest per poter interpretare i risultati ottenuti.

Un progetto che è strettamente legato, molto più di quelli precedentemente menzionati, è l'attività di tesi di K.Leto. In particolare nel progetto che esporrò nelle sezioni seguenti riutilizzo proprio la struttura ad autoencoder più layer di classificazione utilizzate nella tesi di K. Leto [8]. Difatti questo lavoro è il punto di partenza dell'attività che ho svolto durante il tirocinio e di questa tesi.

L'approfondimento che è stato effettuato durante il mio tirocinio, riguardava proprio l'applicazione del modello di K.Leto a un dataset più ampio a livello temporale e su più nodi del supercomputer MARCONI. Inoltre applicare tecniche di oversampling per ridurre il bias della rete neurale verso i record non anomali.

Capitolo 3

Progetto

In questo capitolo si entra nel vivo dell'attività di tesi, ovvero lo sviluppo e l'analisi delle performance del sistema di anomaly detection con l'obiettivo di non solo riuscire a rilevare l'anomalia mentre è in corso, ma riuscire a rilevare pattern nel dataset che precedono l'anomalia in questione.

Data la struttura di Marconi, formata da moltissimi nodi, è stato deciso di prendere come campione 20 nodi e analizzare le loro metriche di performance raccolte dal sistema Examon nei mesi di Gennaio e Maggio.

Alcuni nodi sono stati presi casualmente all'interno del set di nodi disponibili (circa 10). I rimanenti invece sono stati presi con indici vicini a quelli presi casualmente, poiché abbiamo supposto che siano anche fisicamente vicini. Per esempio, il nome di un nodo può essere 'r174c02s01' e come nodo vicino è stato preso 'r174c02s02'.

3.1 Autoencoder semisupervised

Il modello di cui si vuole testare l'efficacia sui dataset descritti precedentemente è un autoencoder allenato in modo semisupervised.

Il modo in cui vogliamo utilizzarlo non è per effettuare una dimensionality reduction del dataset in questione, per il quale basterebbe un approccio unsupervised, ma per un allenamento semi-supervised. Quindi fornire alla rete in fase di training solamente istanze non anomale, in modo tale da avere un autoencoder in grado di saper ricostruire i record non anomali e (teoricamente) incapace di ricostruire record anomali.

Quello che in realtà verrà considerato come output non sarà il record ricostruito, ma appunto la differenza al quadrato o assoluta (MSE o MAE) tra le features del record originale e quello ricostruito, ovvero l'errore di ricostruzione. Questo errore dovrebbe mostrare valori elevati all'incontro di

record che presentano pattern non ricostruibili dall'autoencoder perché non incontrati in fase di training.

I layer utilizzati per questa rete sono 10 relativamente composti da 100, 80, 60, 40 neuroni e infine uno strato che rappresenta lo spazio latente di 20 neuroni per la parte di encoding. La parte di decoding è relativamente composta da i 4 layer elencati prima posti in maniera inversa, quindi 40,60,80,100 e infine l'output layer che avrà la stessa cardinalità delle features presenti nel dataset.

Per quanto riguarda le funzione di attivazione è stato sempre stata usata ReLU a parte per l'output della rete di decoding che è stato usato linear.

Nei risultati riportati nelle sezioni seguenti è stata effettuata una fase di training differente per ogni nodo con il dataset relativo a quello stesso nodo.

3.1.1 Costruzione di train e test set

Il primo problema da risolvere è definire e costruire un modo su come creare training e test set poiché il training set deve essere privo di istanze anomale, quindi la classica funzione "train_test_split" del package scikit-learn non è direttamente utilizzabile, anche perché noi supponiamo anche che l'anomalia in realtà inizi in un periodo temporale precedente alla segnalazione che avviene da NAGIOS, per cui sarebbe ottimo rimuovere dal training set record vicini all'anomalia, sia prima che dopo.

Oltre ai punti vicini all'anomalia, occorre inserire all'interno del test set anche punti random del dataset, al fine di non inserire un bias nel modello.

L'algoritmo creato per definire training set e test set lavora nel modo seguente:

1. Definizione della percentuale del dataset da usare come test set, definire il numero di record da prendere prima e dopo l'anomalia. Ho preferito inserire quest'ultimo parametro in termini di record assoluti poiché ogni record corrisponde a 5 minuti di tempo e inserire questo parametro in termini relativi, in caso di dataset grandi, causerebbe la presenza nel test set di record molto lontani, in termini temporali, dall'anomalia.
2. Estrazione delle anomalie e dei record vicini ad esse. In questa fase potrebbe sussistere il problema di prendere in considerazione lo stesso record più volte in caso di anomalie differenti molto vicine a causa del dover prendere record vicini all'anomalia. Il problema è stato risolto tenendo come variabile l'indice dell'ultimo record preso dall'anomalia precedente e prendendo il massimo tra quest'ultimo indice e il primo record da prendere per l'anomalia successiva.

3. Estrazione dei punti random. Questa fase deve assolutamente essere fatta successivamente all'estrazione dei punti anomali ed eseguita su un set privo dei punti estratti in precedenza per evitare di prendere gli stessi record più volte.
4. Merge dei punti random e delle anomalie.
5. Sottrazione dall'intero dataset dei record random e delle anomalie al fine di creare il training set.

Gli esperimenti sono stati realizzati con i seguenti parametri riguardanti la creazione del test e training set:

- 30% dell'intero dataset come test set
- 24 record (che corrispondono circa a 2h di dati, in caso non siano presenti buchi) da prendere prima e dopo le anomalie

Come parametri per la fase di training dei modelli è stato utilizzato RMSProp come optimizer e come loss il mean squared error.

3.2 Analisi dei grafici

In questa sezione andrò a riportare i grafici nei quali mostrerò i risultati del processo di training di AE-semisupervised confrontandolo con AE-supervised descritto nella sezione Related works.

Dato l'utilizzo di AE semi-supervised, possiamo estrarre l'errore di ricostruzione per ogni feature che non è direttamente graficabile (se non effettuando la media), quindi si è utilizzato un tipo di plot differente da quelli classici chiamato heatmap che consente di identificare in quali periodi temporali e su quali features l'errore di ricostruzione è più marcato.

Come errore è stato scelto MSE su ogni feature da rappresentare, è stata fatta un'analisi anche sul considerare l'errore massimo tra le feature. Questo porta ad avere un segnale di errore sensibile solo a una singola feature (quella con errore massimo) e non a tutte quelle disponibili, e se è presente un errore di ricostruzione permanente su una feature con un errore alto potrebbe portare a dei plot non significativi.

3.2.1 Analisi di Gennaio

r129c12s02

Nella figura 3.1 si può vedere che l'AE semi-supervised ha uno spike di errore prima dell'anomalia stessa 1h prima, anche l'output dell'AE-classifier cambia prima che inizi l'anomalia però senza raggiungere la soglia di 0.5 per dare come output "record anomalo".

L'anomalia dura da 16/01 alle 10:20 fino alle 11:00 e lo spike si ha alle 08:50. Inoltre è possibile evidenziare un buco temporale nei dati (che sarà presente in tutti gli altri nodi in questo periodo) che va da dalle 11:00 alle 18:25 del giorno stesso.

La relativa heatmap 3.2 nella quale è possibile notare come al record 15 (16/01 08:50) sia presente un record con molti errori di ricostruzione, al quale corrisponde il nostro spike.

Durante il mese di Gennaio è stata eseguita una manutenzione durante il giorno 23/01 che è stato eliminato dal dataset poiché le operazioni di manutenzione non dovrebbero rappresentare il comportamento a regime del sistema. Anche dopo la fine della manutenzione è però riportata la label di anomalia durante il giorno 24/01 nel quale possiamo identificare molti errori di ricostruzione, uno in particolare "avg:boottime" mi ha incuriosito e ho indagato sulla sua natura.

La metrica boottime indica in Unix epoch la data di boot del nodo e convertendo il valore presente nel dataset si può rilevare che durante la ma-

avg:SysBrd_3_3V	0.958113
avg:load_fifteen	0.606478
avg:load_five	0.339744
var:load_five	0.286660
var:load_fifteen	0.200711
avg:DC_Energy	0.189542
var:proc_total	0.131108
avg:mem_free	0.131018
var:load_one	0.116976
avg:Sys_Power	0.109989

Tabella 3.1: Top 10 errori di r129c12s02 durante l'anomalia del 16/01/2020

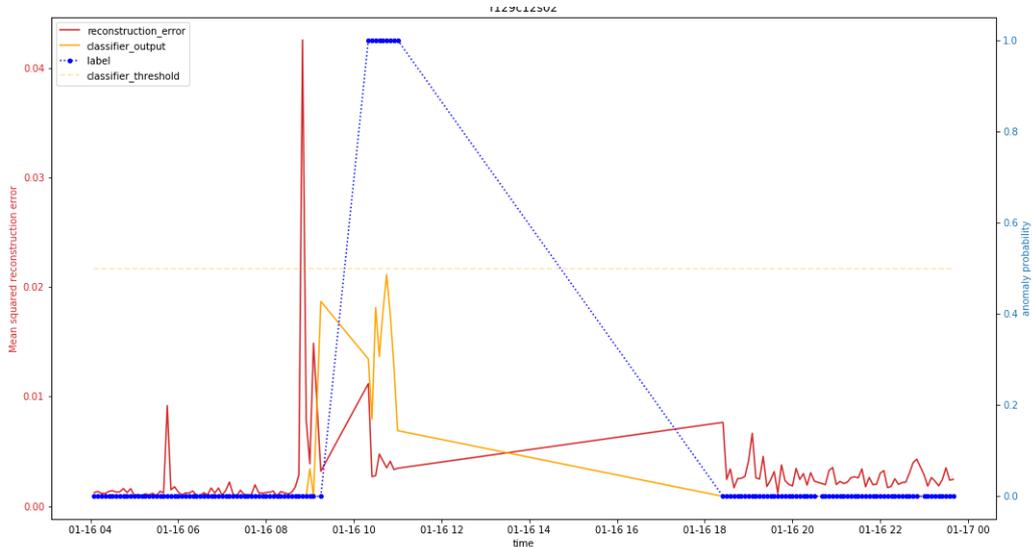


Figura 3.1: Comparazione AE-supervised AE-semisupervised r129c12s02 16 Gennaio

nutenzione il nodo è stato riavviato portando perciò a un cambiamento del valore di quest'ultima. Essendo questa metrica slegata dalle metriche di performance del nodo e non potendo fornire una vera e propria informazione verrà ignorata nei futuri training, inoltre l'anomalia segnalata durante il periodo di manutenzione verrà ignorata nell'analisi.

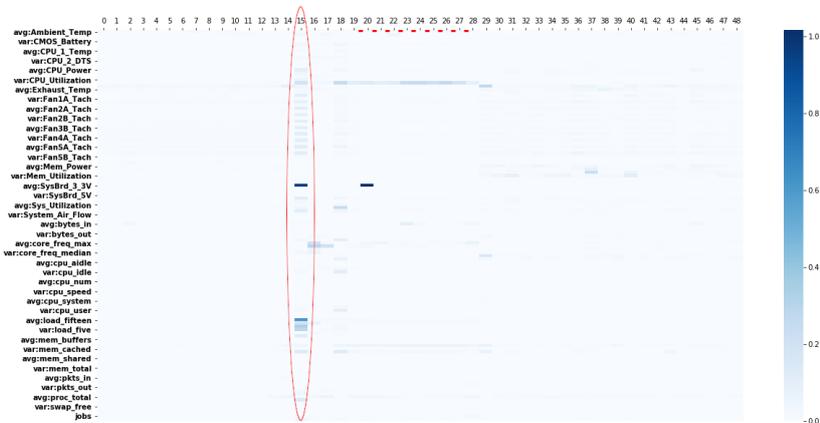


Figura 3.2: Heatmap r129c12s02 16 gennaio



Figura 3.3: Heatmap r129c12s02 24 Gennaio

r136c03s02

Il training su questo dataset ha fatto emergere un comportamento molto particolare, ovvero la presenza di un errore di ricostruzione permanente su una singola feature ovvero "avg:SysBrd_3_3V", questo errore di ricostruzione è presente nell'intero dataset.

La durata del periodo anomalo è di 25 minuti, dalle 10:25 alle 10:45, l'errore medio durante il periodo anomalo è riportato in tabella 3.2.

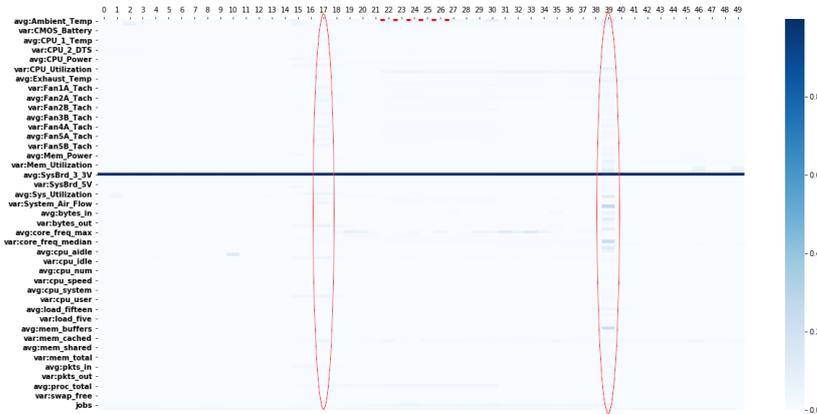


Figura 3.4: Heatmap 136c03s02 16 Gennaio

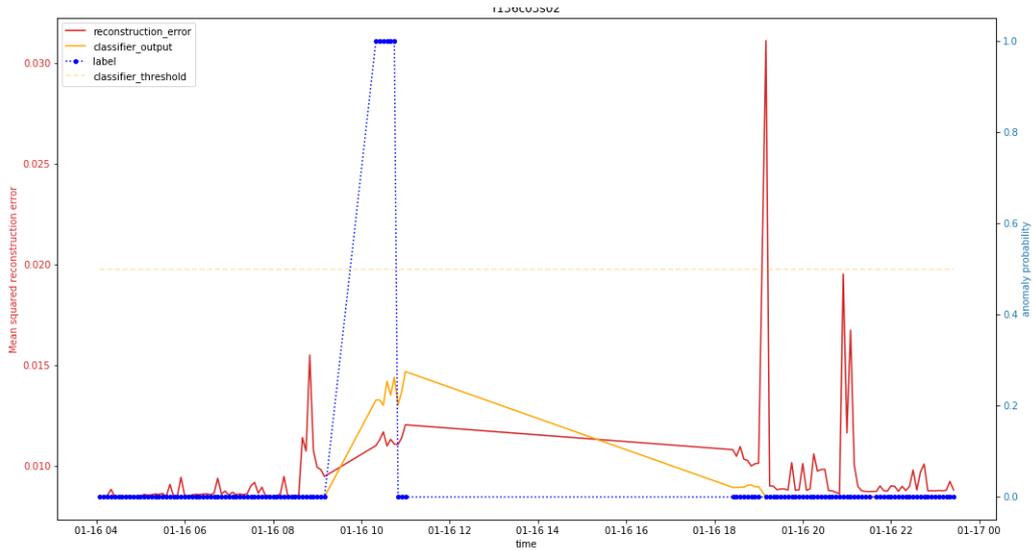


Figura 3.5: Comparazione AE-supervised AE-semisupervised r136c03s02 gennaio

Nell'heatmap 3.4 ho evidenziato i punti ai quali fanno riferimento i due spike in 3.5, prima e dopo l'anomalia. Come si può vedere anche la presenza di un errore di ricostruzione permanente con valore molto alto rispetto agli altri può essere mitigato tramite l'utilizzo dell'errore medio.

AE-supervised in questo caso non ha variazioni nell'output nei periodi precedenti all'anomalia e anche durante il periodo anomalo stesso nonostante

avg:SysBrd_3_3V	0.991673
avg:DC_Energy	0.041920
jobs	0.036308
avg:Ambient_Temp	0.034549
avg:proc_total	0.031750
avg:mem_free	0.018234
avg:Fan5A_Tach	0.016678
var:Ambient_Temp	0.015240
avg:Fan1A_Tach	0.011807
mean_error	0.011252

Tabella 3.2: Errore medio durante il periodo anomalo 16/01 di r136c03s02

abbia un variazione sensibile non riesce a superare la soglia di 0.5.

r140c12s04

Anche in questo nodo è possibile trovare un errore permanente, che in realtà è presente in 3/4 del dataset sulla feature 'avg:mem_buffers'. L'anomalia inizia alle 10:20 e finisce alle 10:45.

Nel plot 3.6 è possibile notare 1 spike poco prima dell'anomalia e uno durante la stessa che sono stati evidenziati in rosso in 3.7.

Poco prima di un altro spike successivo all'anomalia è possibile evidenziare due errori (segnati in arancione nella heatmap) su una due feature che continuano per 7 record (35 minuti) sulla metrica "avg:CMOS_Battery" e "avg:proc_total".

In 3.6 è possibile vedere come durante l'anomalia AE-supervised riesca a superare la soglia 0.5 in alcuni punti, anche dopo la fine dell'anomalia stessa. Si può anche identificare che un comportamento sospetto dopo l'anomalia alle 19:10 (che corrisponde al record 41 evidenziato in rosso in 3.7) viene rilevato anche da AE-supervised.

In particolare nel record 41 gli errori sono concentrati in: avg:mem_buffers, avg:CMOS_Battery (la stessa che presenta errore nei record precedenti) e var:proc_total.

Il nodo r140c12s04 presenta una seconda anomalia durante il periodo di Gennaio, che inizia il 17 Gennaio alle 12:35 e termina lo stesso giorno alle 18:30. Non essendo presenti buchi in questo periodo temporale ho diviso le heatmap dell'errore in tre parti, in modo da poter effettuare uno zoom sull'inizio dell'anomalia, sulla parte centrale e sulla parte finale.

In 3.8 è possibile notare che già da circa le 11 dello stesso giorno siano presenti spike nell'errore medio di ricostruzione delle features, molto alto poco prima che inizi (12:30).

avg:mem_buffers	0.982214
avg:SysBrd_3_3V	0.159411
avg:proc_total	0.023720
avg:core_freq_max	0.021220
avg:CMOS_Battery	0.020571
avg:SysBrd_12V	0.017637
avg:mem_cached	0.017506
var:SysBrd_12V	0.014194
mean_error	0.010940
avg:core_freq_median	0.007104

Tabella 3.3: Errore medio durante il periodo Anomalo di r140c12s04

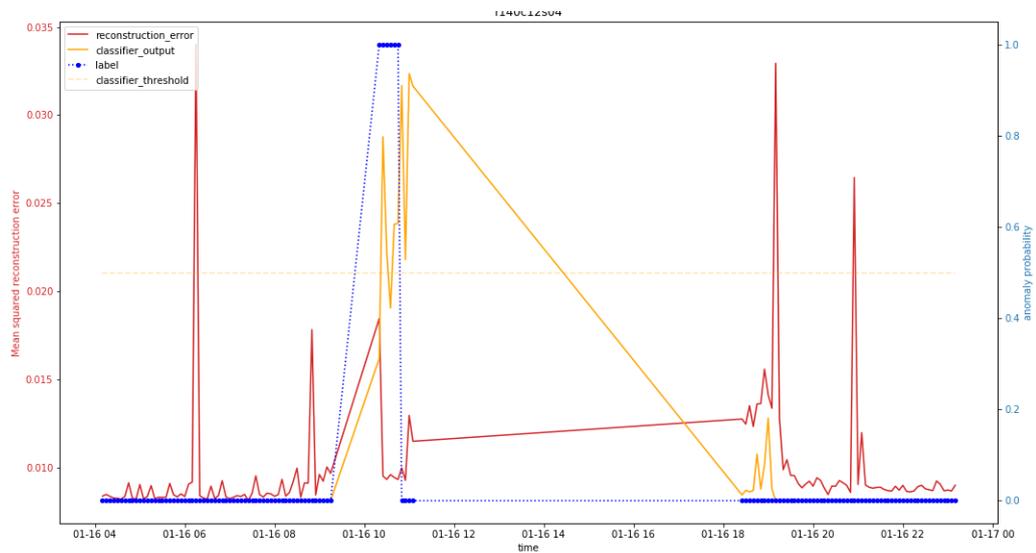


Figura 3.6: Comparazione AE-supervised AE-semisupervised r140c120s04 16 Gennaio

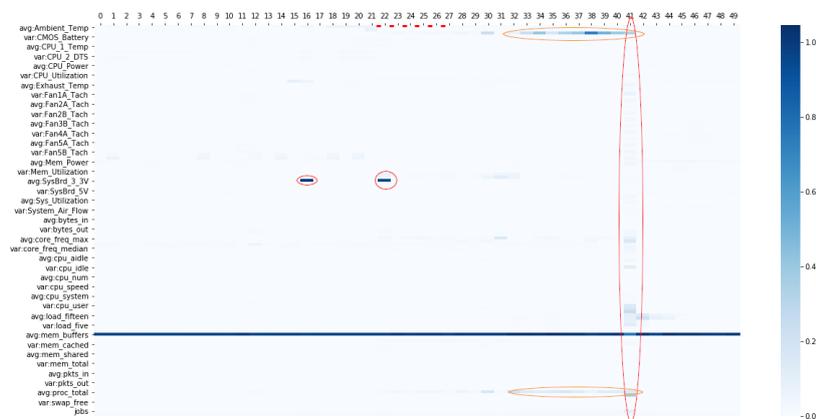


Figura 3.7: Heatmap r140c12s02 16 Gennaio

In 3.9 ho evidenziato la posizione degli spike nel quale è possibile identificare che oltre all'errore permanente su "avg:mem_buffers" è presente un errore che permane per tutto l'inizio dell'anomalia, nei 2 record precedenti ed è già presente nel primo spike ed è sulla feature "avg:CMOS_Battery".

In quest'anomalia AE-supervised identifica dopo due record l'anomalia con output molto alto che rimane sempre sopra la soglia dello 0.5, riuscendo

quindi a identificarla con successo.

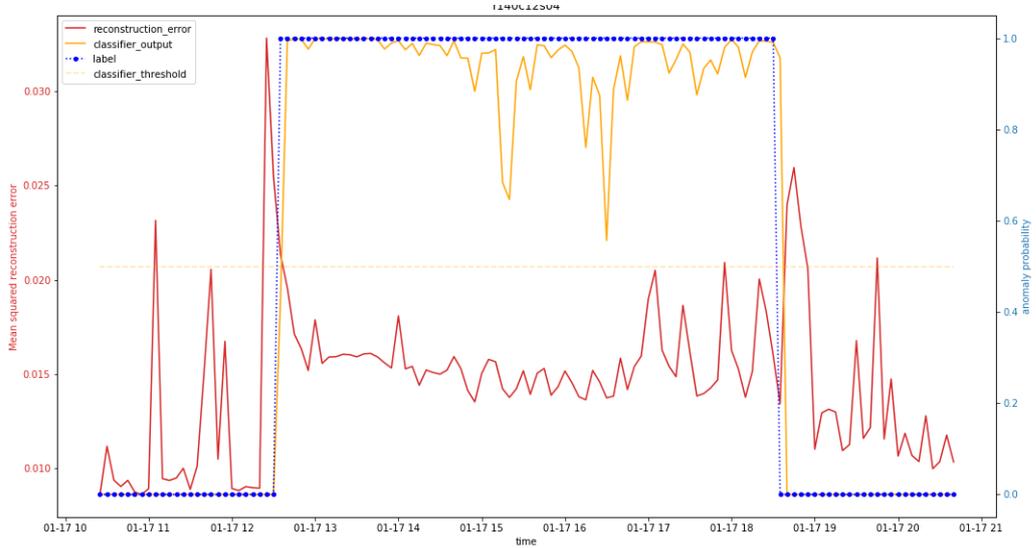


Figura 3.8: Comparazione AE-supervised AE-semisupervised r140c120s04 17 Gennaio

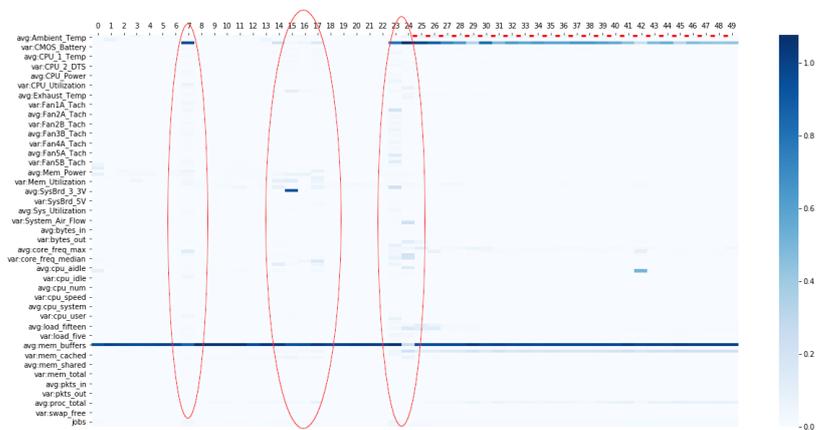


Figura 3.9: Heatmap r140c120s04 17 Gennaio - Inizio anomalia

Durante la parte centrale e la fine dell'anomalia è costantemente presente l'errore su "avg:CMOS_Battery" che si dirada poco dopo la fine dell'anomalia.



Figura 3.10: Heatmap r140c120s04 17 Gennaio - Centro anomalia

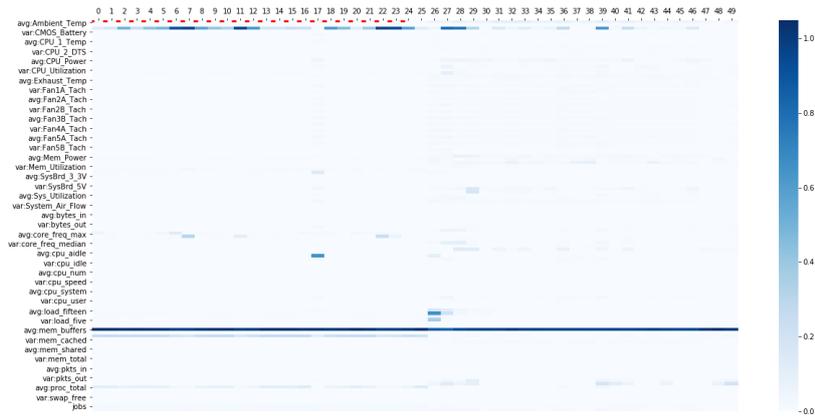


Figura 3.11: Heatmap r140c120s04 17 Gennaio - Fine anomalia

r174c02s02

Anche in questo nodo sono presenti due anomalie, la prima è quella che accomuna anche i nodi precedenti del 16 Gennaio.

L'anomalia in questione dura dalle 10:20 alle 10:45, si hanno molti spike nell'errore, che si accentuano poco prima che l'anomalia inizi, anche AE-supervised ha un picco poco prima dell'anomalia. Nei periodi successivi però l'errore è molto basso, quindi si può attribuire questi spike all'anomalia.

avg:CMOS_Battery	0.238220
avg:core_freq_max	0.077955
avg:mem_cached	0.059545
jobs	0.041672
avg:Ambient_Temp	0.039232
avg:bytes_in	0.024394
avg:proc_total	0.022378
avg:load_fifteen	0.016327
avg:Exhaust_Temp	0.016040
var:core_freq_max	0.013711

Tabella 3.4: Errore medio durante il periodo anomalo 16/01 di r174c02s02

Nell'heatmap 3.13 si può notare come prima dell'anomalia siano già presenti molti errori sparsi su più features. Durante l'anomalia però è evidente l'errore su una feature in particolare "avg:CMOS_Battery". Questo errore continua anche nei record successivi, in particolare fino al record 30 (16/01 11:00).

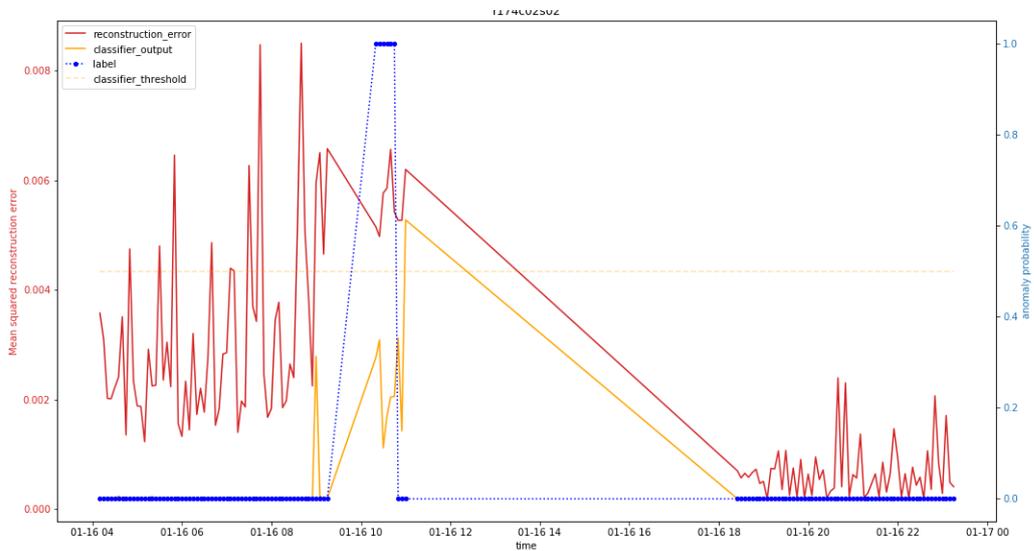


Figura 3.12: Comparazione AE-supervised AE-semisupervised r174c02s02 16 Gennaio



Figura 3.13: Heatmap r174c02s02 16 Gennaio

L'anomalia successiva riguarda il giorno 15 Gennaio, l'anomalia inizia alle 12:05 e termina lo stesso giorno alle 15:00.

Nel grafico 3.14 si può vedere come dalle 10:00 inizino a essere presenti picchi nell'errore e poco dopo le 11:00 anche l'AE-supervised ha un output più alto ma comunque minore di 0.5. Durante l'anomalia invece AE-supervised riesce a superare la soglia di 0.5 quasi costantemente mentre l'errore si abbassa drasticamente dai valori visti nei picchi.

Nell'heatmap 3.15 è possibile notare come nei record precedenti all'inizio dell'anomalia siano presenti record con errori di ricostruzione su più features. In comune tra l'anomalia e i record precedenti si ha 'avg:CMOS_Battery' ormai incontrato più volte durante l'analisi.

avg:CMOS_Battery	0.121175
avg:proc_total	0.093894
avg:mem_cached	0.084152
avg:core_freq_max	0.046662
var:SysBrd_12V	0.031698
jobs	0.028256
avg:SysBrd_12V	0.023216
avg:load_fifteen	0.022738
var:core_freq_max	0.022497
avg:DC_Energy	0.014784

Tabella 3.5: Errore medio durante l'anomalia del 15 Gennaio r174c02s02

avg:load_five	0.442556
avg:load_fifteen	0.342488
avg:CMOS_Battery	0.165045
var:CPU_2_Temp	0.122788
avg:load_one	0.107330
var:CPU_1_Temp	0.105800
avg:mem_free	0.104432
avg:mem_cached	0.101760
var:System_Air_Flow	0.090787
avg:SysBrd_12V	0.077290

Tabella 3.6: Errore medio dei record precedenti all'anomalia del 15 Gennaio di r174c02s02

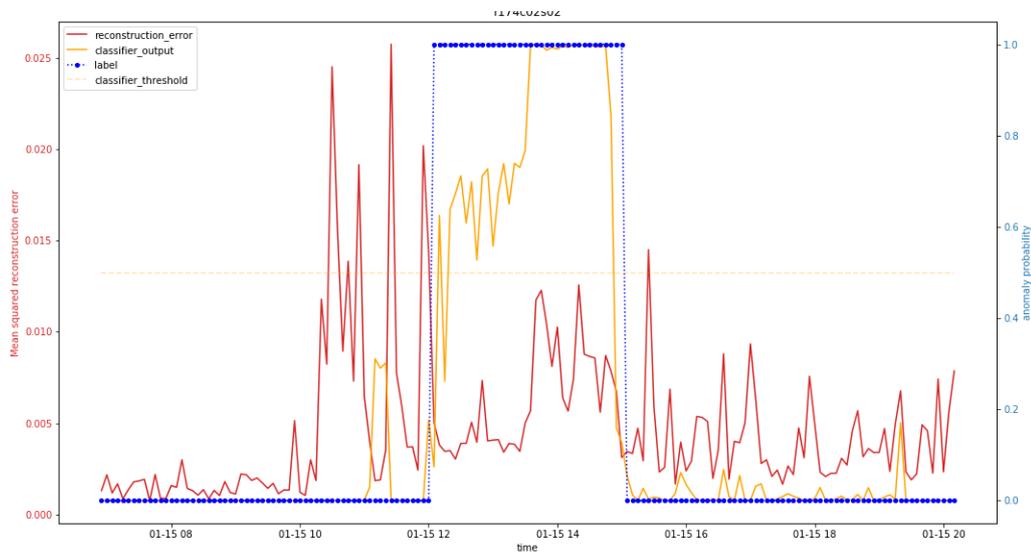


Figura 3.14: Comparazione AE-supervised AE-semisupervised r174c02s02 15 Gennaio

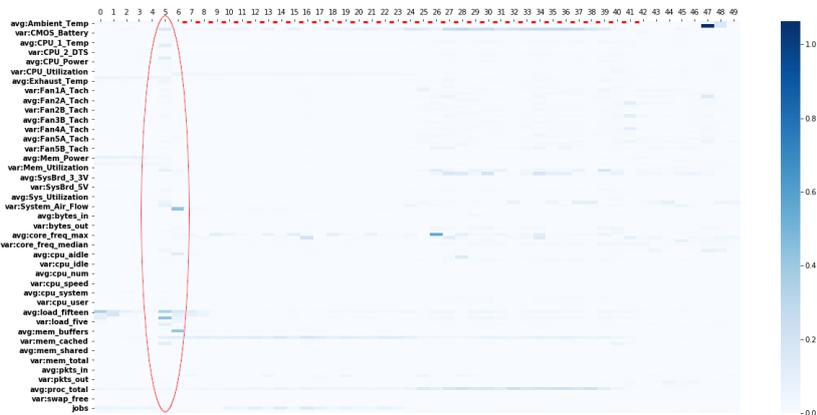


Figura 3.15: Heatmap r174c2s02 15 Gennaio

r173c10s01

Questo nodo presenta l'anomalia del 16 Gennaio dalle 10:20 alle 10:45. Si può notare in 3.16 come siano presenti picchi significativi prima dell'inizio dell'anomalia. In questo caso l'AE-supervised non riesce a rilevare l'anomalia in anticipo.

Nella figura 3.2 si può notare come gli errori siano presenti solo poco

jobs	0.292900
avg:SysBrd_3_3V	0.106887
avg:mem_cached	0.037543
avg:CMOS_Battery	0.030886
avg:core_freq_max	0.025985
avg:load_fifteen	0.019689
avg:Ambient_Temp	0.010582
avg:boottime	0.006195
var:core_freq_max	0.005957
mean_error	0.004791

Tabella 3.7: Errore medio durante l'anomalia del 16 Gennaio r173c10s01

prima e durante l'anomalia.

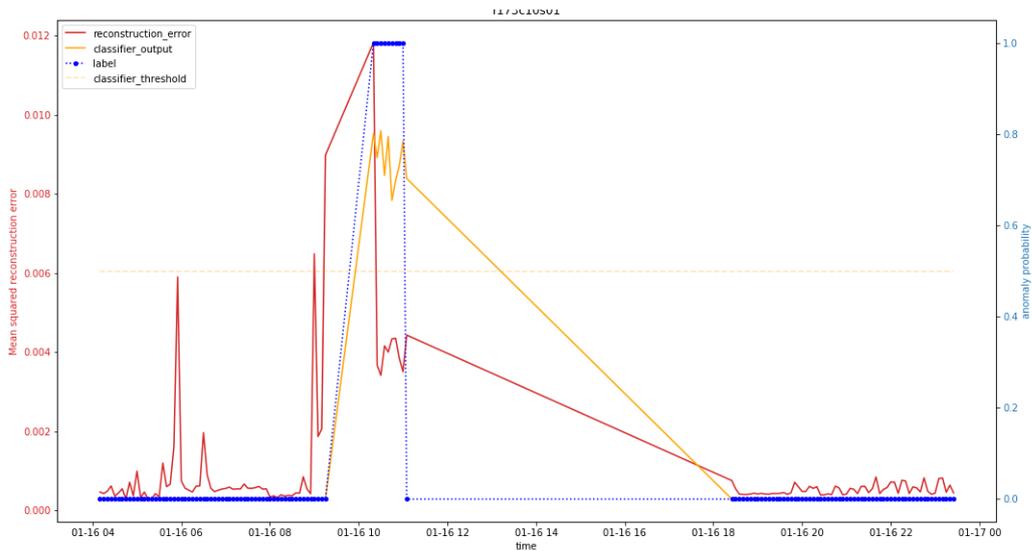


Figura 3.16: Comparazione AE-supervised AE-semisupervised r173c10s01 16 Gennaio

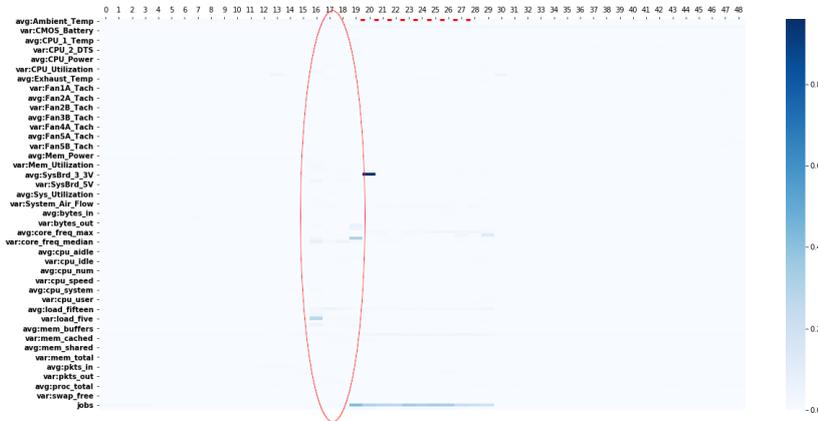


Figura 3.17: Heatmap r173c10s01 16 Gennaio

r162c05s02

Classica anomalia del 16 Gennaio dalle 10:20 alle 10:45, si può osservare in 3.18 un picco netto dell'errore prima che inizi l'anomalia e che l'AE supervised non riesce a identificare nulla prima che l'anomalia incominci, ma durante l'anomalia è praticamente sempre sopra la soglia di 0.5, riuscendo quindi ad identificare l'anomalia correttamente ma senza effettuare la predizione in anticipo.

È inoltre possibile identificare il picco in 3.19 cerchiato in rosso, nella tabella 3.9 è presente da cosa è composto l'errore medio prima che l'anomalia accada e in 3.8 l'errore durante l'anomalia.

avg:mem_cached	0.040586
avg:CMOS_Battery	0.025231
var:core_freq_max	0.018104
avg:core_freq_max	0.009117
avg:proc_total	0.008921
var:Ambient_Temp	0.008881
jobs	0.007149
avg:load_fifteen	0.004436
avg:Ambient_Temp	0.003501
avg:bytes_in	0.001908

Tabella 3.8: Errore medio durante il picco di r162c05s02 durante l'anomalia del 16 Gennaio

var:cpu_idle	0.318270
var:cpu_user	0.316601
avg:Mem_Power	0.142981
var:CPU_Utilization	0.142178
avg:proc_total	0.103711
var:load_one	0.072145
avg:Sys_Utilization	0.066047
var:Sys_Utilization	0.060505
avg:Fan5B_Tach	0.056978
avg:cpu_idle	0.055623

Tabella 3.9: Errore medio durante il picco di r162c05s02 prima l'anomalia del 16 Gennaio

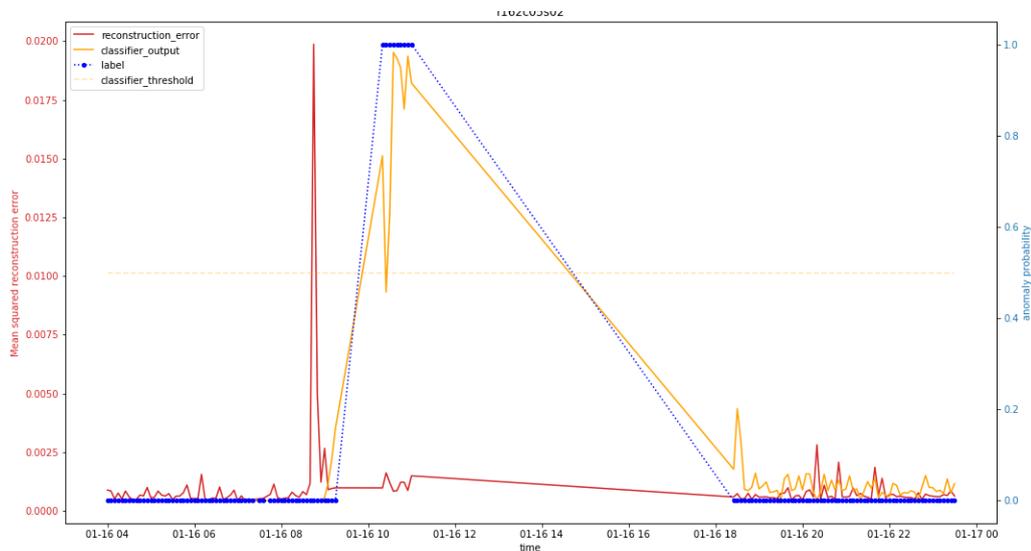


Figura 3.18: Comparazione AE-supervised AE-semisupervised r162c05s02 16 Gennaio

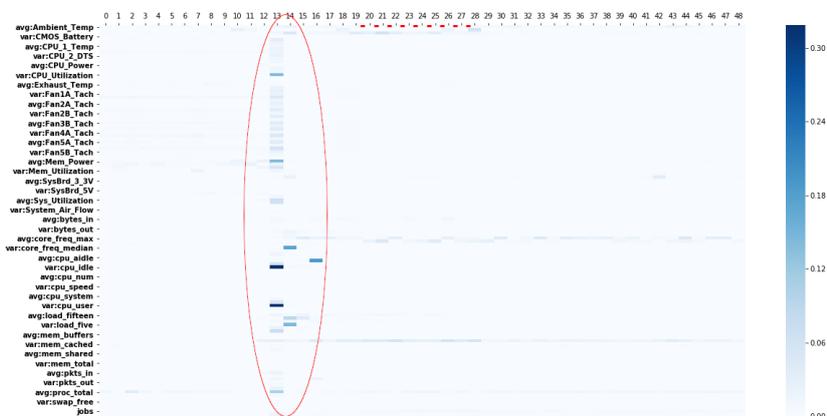


Figura 3.19: Heatmap r162c05s02 16 Gennaio

Il nodo r162c05s02 presenta un'anomalia anche il 17 Gennaio che inizia alle 14:25 e termina alle 18:45, poiché è molto lunga riporterò solo l'inizio e la fine dell'anomalia come heatmap e l'anomalia intera come plot classico.

Nel grafico 3.20 è presente un innalzamento netto dell'errore a partire circa dalle 11:00 e molto alto vicino all'inizio dell'anomalia. Da notare anche che nei periodi precedenti poco dopo le 13 anche l'AE-supervised supera la

soglia di 0.5 dando un falso positivo prima che l'anomalia incominci. Durante l'anomalia l'AE-supervised si comporta bene dando sempre un output maggiore di 0.5, mentre l'AE-semisupervised presenta un errore molto basso se non al termine dell'anomalia.

In figura 3.21 è possibile notare come poco prima dell'anomalia (dalle 13:35 alle 13:50) siano presenti errori su più features riportati nella tabella 3.10. Durante l'anomalia non è possibile vedere nell'heatmap nessun errore evidente.

Riguardo alla fine dell'anomalia, si può notare nell'heatmap 3.22 che proprio al record 24 (ultimo record anomalo) si ha un innalzamento degli errori di ricostruzioni su più feature, lo stesso anche nei record seguenti. Questo fenomeno potrebbe essere attribuito all'aver rimesso il nodo come "available" all'interno di Marconi, questo può essere visto anche dalla tabella 3.11 nel quale l'errore di ricostruzione è concentrato su metriche di rete e carico del nodo.

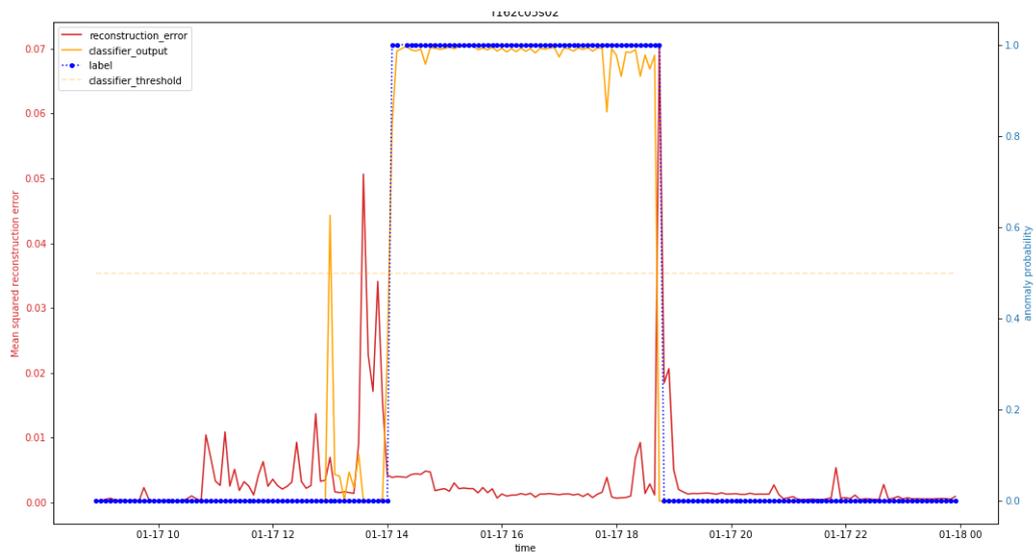


Figura 3.20: Comparazione AE-supervised AE-semisupervised r162c05s02 17 Gennaio

avg:core_freq_min	0.510898
avg:cpu_system	0.293083
avg:mem_cached	0.222006
avg:bytes_in	0.176699
var:cpu_wio	0.172949
avg:cpu_wio	0.155225
var:bytes_in	0.150885
var:core_freq_min	0.136757
avg:load_fifteen	0.104643
var:load_five	0.100549

Tabella 3.10: Errore medio prima dell'anomalia del 17/01 di r162c05s02

var:pkts_out	0.976953
var:bytes_out	0.918993
avg:load_fifteen	0.633210
avg:pkts_out	0.540639
var:load_fifteen	0.497083
var:Mem_Power	0.482770
avg:bytes_out	0.467375
var:Mem_Utilization	0.399120
var:load_five	0.357036
var:Exhaust_Temp	0.327208

Tabella 3.11: Errore medio del record di fine anomalia del 17/01 di r162c05s02

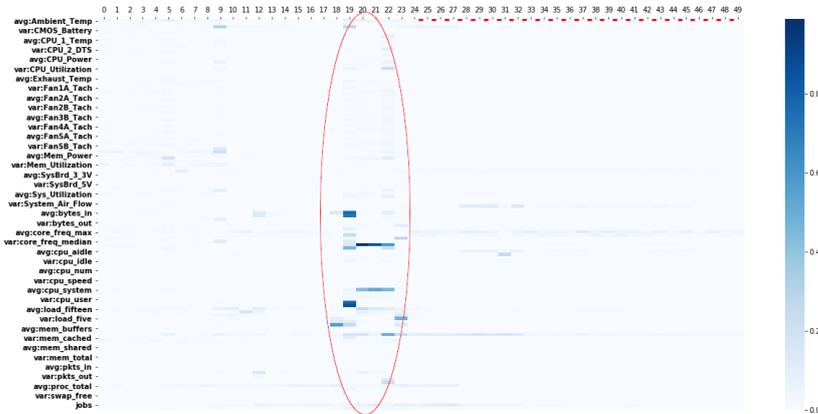


Figura 3.21: Heatmap r162c05s02 17 Gennaio - Inizio Anomalia



Figura 3.22: Heatmap r162c05s02 17 Gennaio - Fine Anomalia

r129c12s03

Questo nodo presenta due anomalie, la prima è quella del 4 Gennaio che dura dalle 14:50 alle 15:00. Sono presenti due record anomali solamente invece di 3 poiché è presente un buco nei dati.

Come è possibile notare in 3.23 sono presenti numerose spike attorno, prima e dopo l'anomalia, viene quindi possibile affermare che l'AE-semisupervised non abbia identificato a dovere l'anomalia, ma nemmeno

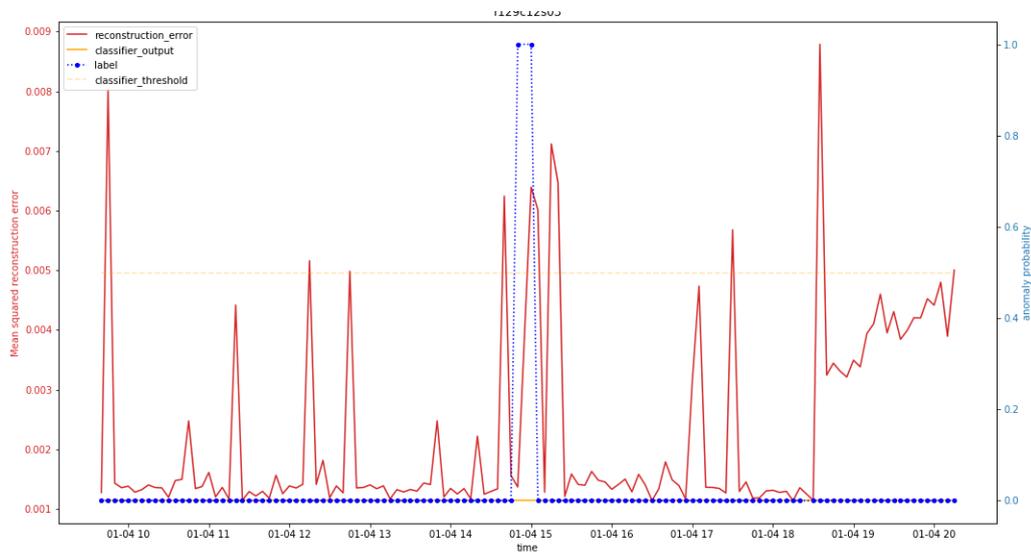


Figura 3.23: Comparazione AE-supervised AE-semisupervised r129c12s03 4 Gennaio

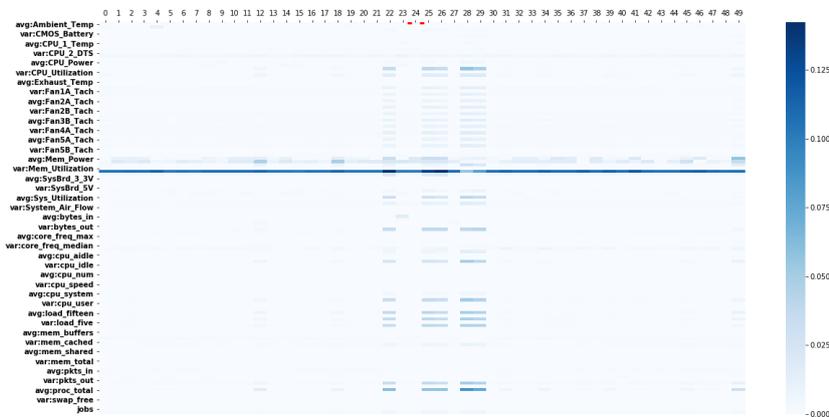


Figura 3.24: Heatmap r129c12s03 4 Gennaio

l'AE-supervised come si può vedere non ha alcuna variazione nel suo output durante il periodo.

Osservando invece l'heatmap 3.24 è possibile notare come nell'intorno dell'anomalia siano presenti errori che durano più di un record sulle stesse features riportate nella tabella 3.12.

avg:SysBrd_12V	0.142226
avg:proc_total	0.060656
avg:load_fifteen	0.036808
avg:load_five	0.036106
avg:CPU_Utilization	0.035273
avg:proc_run	0.034826
avg:load_one	0.034824
avg:Mem_Power	0.034777
avg:core_freq_avg	0.034140
avg:cpu_user	0.032111

Tabella 3.12: Errori frequenti attorno all'anomalia di r129c12s03 4 Gennaio

Anche in questo nodo è presente l'anomalia del 16 Gennaio dalle 10:20 alle 11:00. Come nelle anomalie precedenti poco prima dell'anomalia, circa un ora prima, si ha uno spike nell'errore che è molto più basso nei periodi successivi. AE-supervised riesce a identificare l'anomalia durante tutto il periodo anomalo indicato fornendo un output molto superiore alla soglia di 0.5.

Come si può vedere in 3.26 l'errore è concentrato tra i record 13-18 che fanno riferimento agli istanti temporali (8:40 - 9:05) il quale errore è riportato nella tabella 3.13. Inoltre la sezione evidenziata nell'heatmap si riferisce alla metrica "avg:SysBrd_12V" che ha un errore lungo tutta l'anomalia e il periodo precedente.

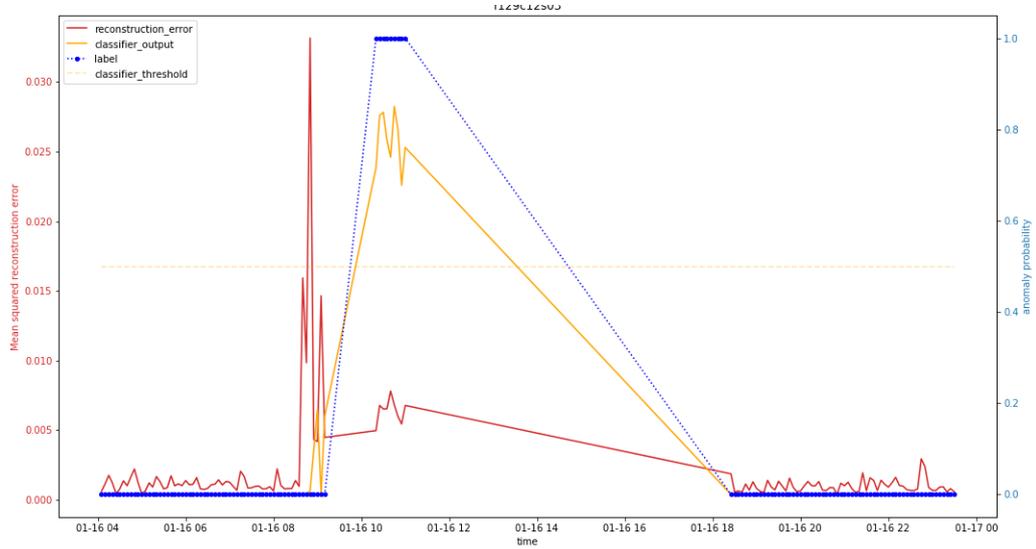


Figura 3.25: Comparazione AE-supervised AE-semisupervised r129c12s03 16 Gennaio

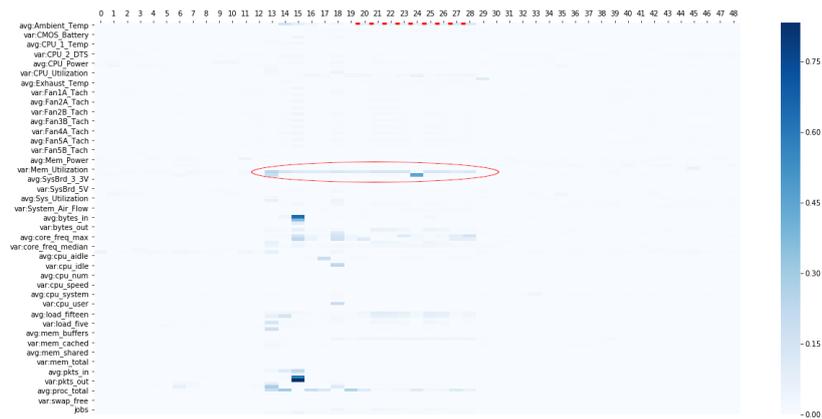


Figura 3.26: Heatmap r129c12s03 16 Gennaio

avg:proc_total	0.154530
avg:SysBrd_12V	0.137894
var:pkts_out	0.118852
avg:bytes_in	0.099152
avg:pkts_out	0.086984
var:core_freq_max	0.068620
avg:core_freq_max	0.058241
var:proc_run	0.054831
var:bytes_in	0.051798
avg:pkts_in	0.049721

Tabella 3.13: Errore medio del periodo precedente all'anomalia del 16 Gennaio di r129c12s03

r141c09s01

Questo nodo ha un'anomalia il 16 Gennaio, ormai molto ricorrente. Dura dalle 10:20 alle 10:45 ed è ormai ricorrente il pattern di picco poco prima dell'anomalia (Figura 3.27) ma non presenta un decadimento dell'errore così evidente come in altri casi. In questo caso AE-supervised non riesce a superare la soglia per dare un output differente, ma si può chiaramente vedere che la rete è indecisa sull'output dando circa un 40% di possibilità che il record sia anomalo.

Nell'heatmap 3.28 è possibile evidenziare una serie di errori che persistono prima dell'anomalia (cerchio rosso in alto) e features che presentano errori durante l'anomalia e nei record immediatamente successivi. Nel periodo precedente l'errore più evidente è quello su "avg:CMOS_Battery" ed errori lievi su metriche che riguardano la ventilazione del nodo. Mentre durante l'anomalia si hanno errori di ricostruzione sulle metriche "avg:proc_total", "jobs", avg:mem_cached".

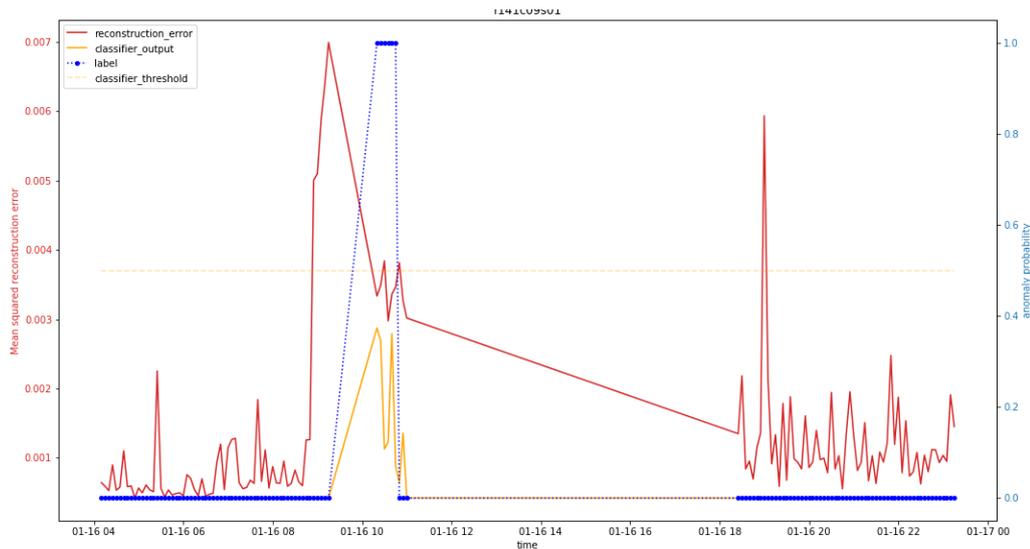


Figura 3.27: Comparazione AE-supervised AE-semisupervised r141c09s01 16 Gennaio



Figura 3.28: Heatmap r141c09s01 16 Gennaio

r135c09s03

Questo nodo presenta una novità riguardo all'anomalia del 16 Gennaio poiché presenta l'anomalia dalle 10:20 alle 11:00, inoltre ne presenta una anche dalle 18:35 alle 19:30 dello stesso giorno.

In figura 3.27 si può notare ormai i classici spike dell'errore prima dell'anomalia delle 10:20 e il successivo calo dell'errore, si nota anche che in questo caso AE-supervised riesce a identificare bene l'anomalia ma non riesce effettuare alcuna rilevazione di quest'ultima in anticipo.

Per quanto riguarda la seconda anomalia si ha un piccolo spike nell'errore di ricostruzione all'inizio e al termine della stessa.

Nell'heatmap 3.30 si può vedere come l'errore di ricostruzione sia molto sparso ed elevato nei momenti precedenti alla prima anomalia dalle 8:45 nel quale l'errore è concentrato sia su media che varianza di queste due feature "Sys_Utilization" e "CPU_Utilization".

La seconda anomalia invece ha come errore dominante quello sulla metrica "avg:mem_cached", errore già presente anche alle 18:25 (record 29 del plot).

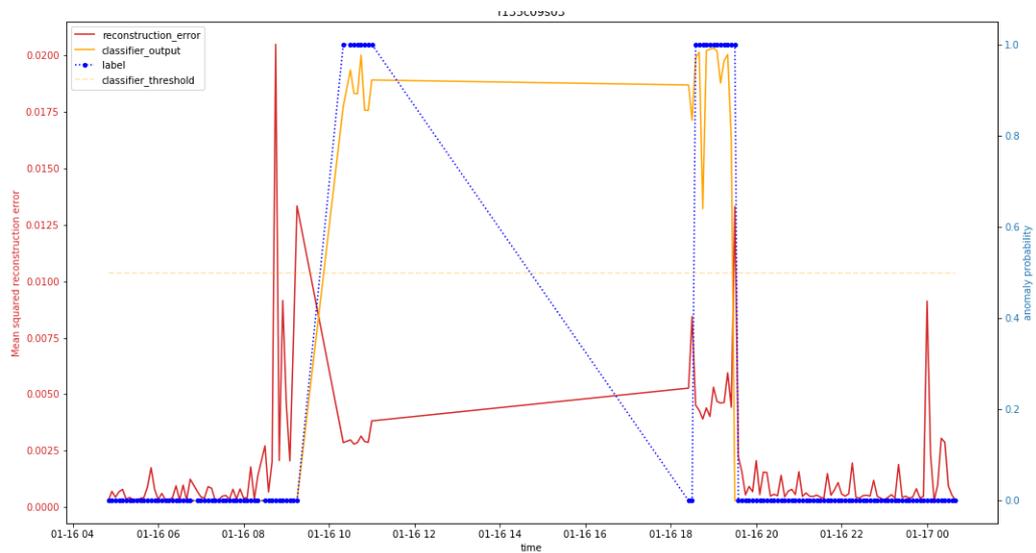


Figura 3.29: Comparazione AE-supervised AE-semisupervised r135c09s03 16 Gennaio

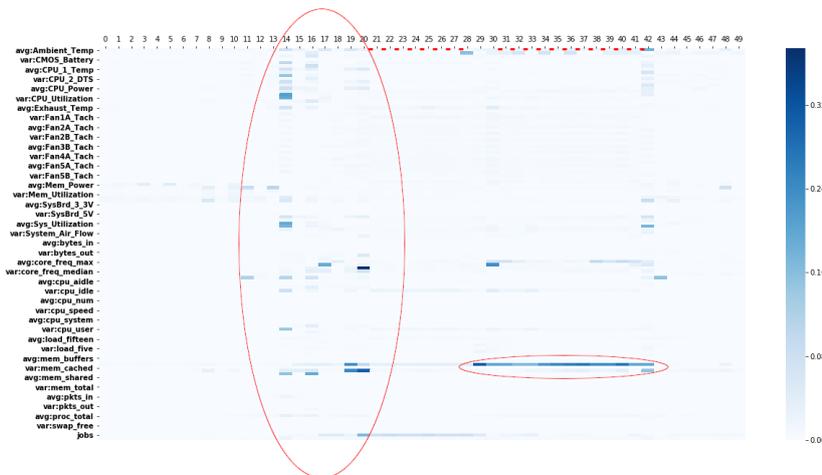


Figura 3.30: Heatmap r135c09s03 16 Gennaio

r162c05s03

Questo nodo presenta due anomalie; il 16 Gennaio e il 17 Gennaio. Inizio con analizzare quella del 16 Gennaio che è quella più frequente all'interno di questo dataset.

In particolare inizia alle 9:05 e termina alle 11:00 col buco di dati dalle 9:15 alle 10:20, quindi inizia circa 1 ora prima rispetto alle altre anomalie

analizzate del 16 Gennaio. È molto evidente il picco, sia nel plot sia nell'heatmap, alle 8:45 quindi poco prima dell'inizio dell'anomalia. L'errore sul record 15 è riportato in tabella 3.14.

Molto interessante anche il comportamento dell'AE-supervised il quale riesce a identificare in modo corretto solo la parte finale dell'anomalia, ma presenta anche variazioni dell'output (in un punto anche maggiore di 0.5) successivamente all'anomalia e il buco di dati.

Durante l'anomalia sebbene abbiano un valore minore agli errori riportati alle 8:45 sono presenti errori continui su 3 metriche in particolare: "jobs", "avg:mem_cached", "avg:DC_Energy". 2

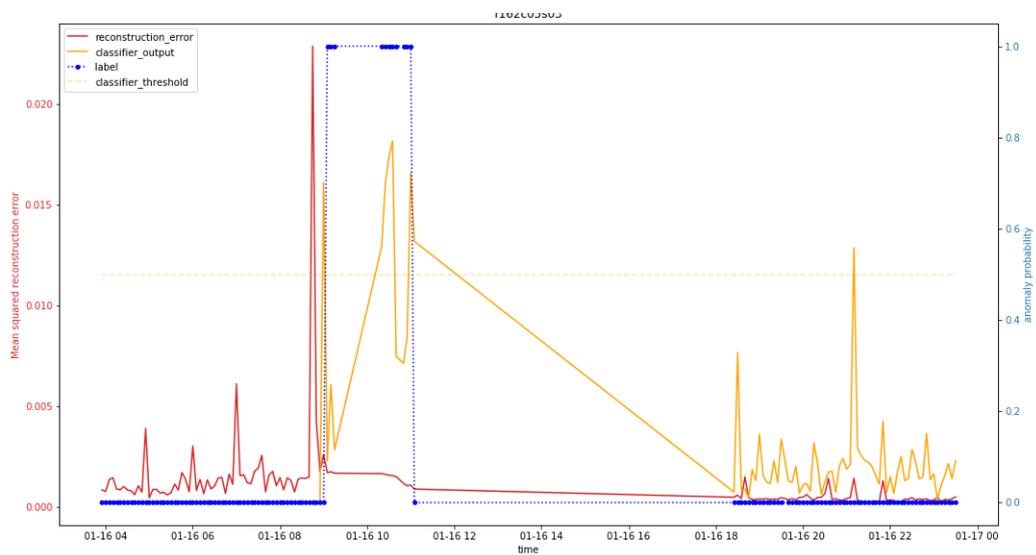


Figura 3.31: Comparazione AE-supervised AE-semisupervised r162c05s03 16 Gennaio

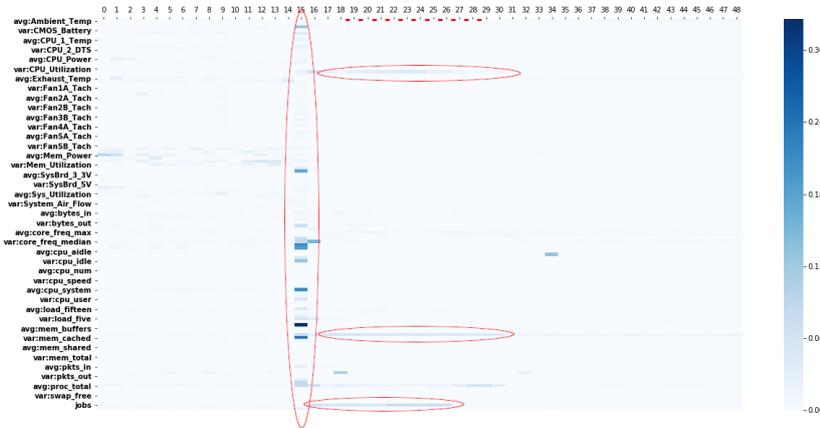


Figura 3.32: Heatmap r162c05s03 16 Gennaio

var:load_one	0.325636
var:mem_cached	0.235832
avg:cpu_system	0.211893
avg:core_freq_min	0.208688
var:core_freq_min	0.190265
var:SysBrd_12V	0.168416
avg:CMOS_Battery	0.128366
var:cpu_idle	0.127143
var:proc_run	0.087592
var:core_freq_median	0.082239

Tabella 3.14: Errore delle 8:45 anomalia del 16 Gennaio di r162c05s03

L'anomalia del 17 Gennaio dura dalle 14:05 fino alle 18:45, essendo molto lunga riporterò le heatmap dell'inizio e della fine dell'anomalia.

Nel plot 3.33 è possibile identificare alcuni picchi prima che inizi l'anomalia e anche alcuni appena finita l'anomalia. L'AE-supervised presenta alcune oscillazioni seppur con bassi valori anche prima dell'anomalia, durante l'anomalia ha un cambiamento dell'output e riesce anche a superare la soglia dello 0.5 che gli permette di identificare la soglia in modo corretto. Nell'heatmap 3.34, ho evidenziato i picchi e la metrica che presenta un errore durante tutta l'anomalia ovvero "avg:mem_cached".

Dopo la fine dell'anomalia sono presenti errori riguardanti la cpu e le metriche di carico del nodo, fattore che può essere associato alla transizione

da idle a "running".

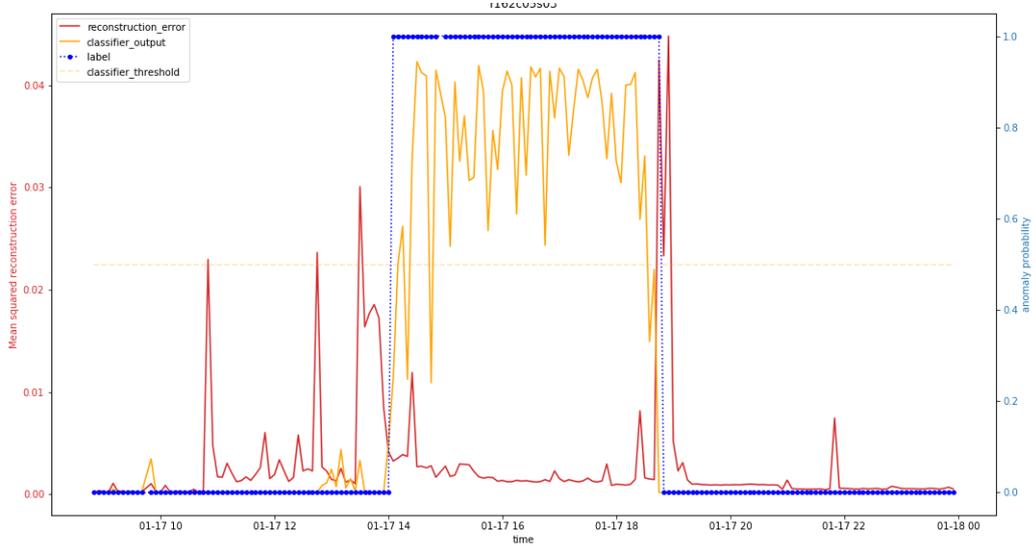


Figura 3.33: Comparazione AE-supervised AE-semisupervised r162c05s03 17 Gennaio

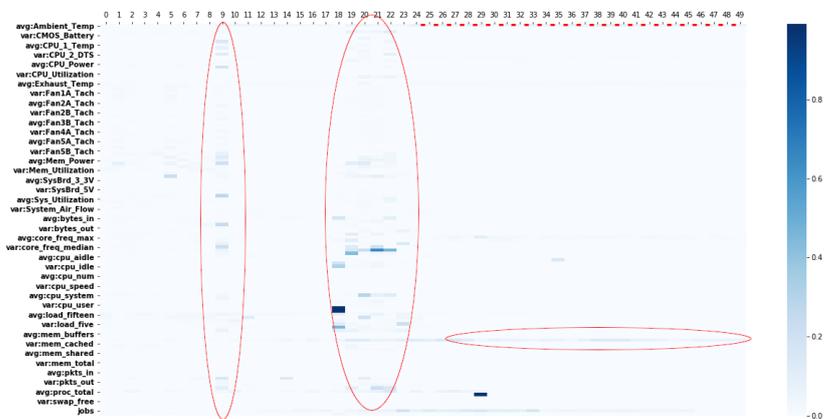


Figura 3.34: Heatmap r162c05s03 17 Gennaio - Inizio



Figura 3.35: Heatmap r162c05s03 17 Gennaio - Fine

r136c04s04

Questo nodo presenta le due anomalie del 16 Gennaio una dalle 10:20 alle 11:00 e la seconda dalle 18:35 alle 19:30. Analizzo la prima anomalia con la quale ormai sono familiare.

In questo caso sono comunque presenti picchi dell'errore di ricostruzione prima che l'anomalia incominci ma l'errore non cala una volta che l'anomalia è iniziata, l'AE-supervised identifica subito correttamente l'anomalia con un output molto vicino 1 (il massimo possibile), anche durante la seconda anomalia. È molto interessante vedere come alla fine della seconda anomalia sia AE-supervised che semi-supervised mostrino un comportamento oscillatorio con errori e output molto più alti rispetto all'inizio della prima anomalia.

Nell'heatmap 3.37 è possibile notare come una metrica in particolare abbia errore durante i periodi anomali e anche tra essi (nonostante sia presente un buco di dati dal record 28 al record 29 (dalle 11 alle 18:25), la metrica in questione è "avg:CMOS_Battery".

Molto particolare il record 29 dell'heatmap che fa riferimento all'istante di tempo delle 18:25 durante il quale sono presenti errori molto marcati mostrati in tabella 3.15.

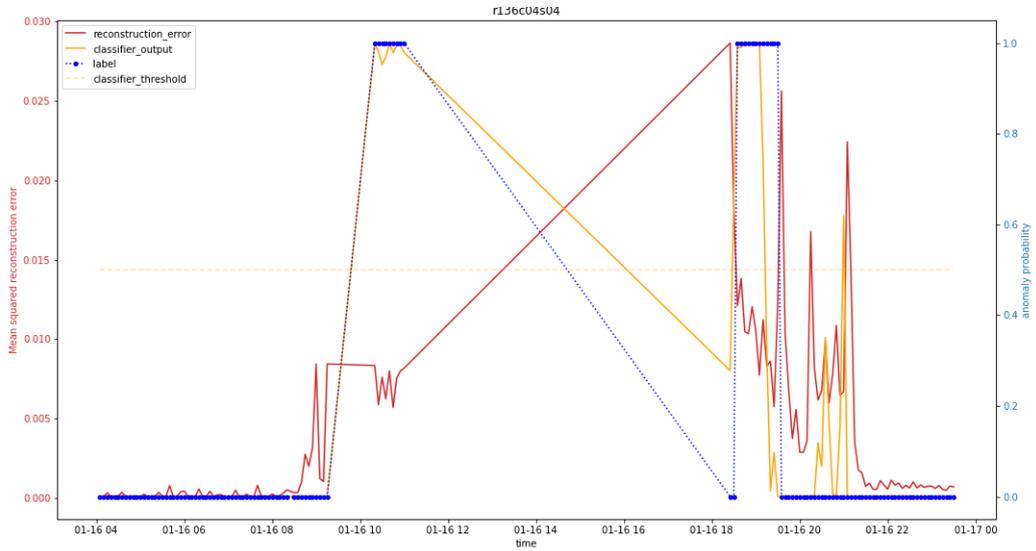


Figura 3.36: Comparazione AE-supervised AE-semisupervised 136c04s04 16 Gennaio

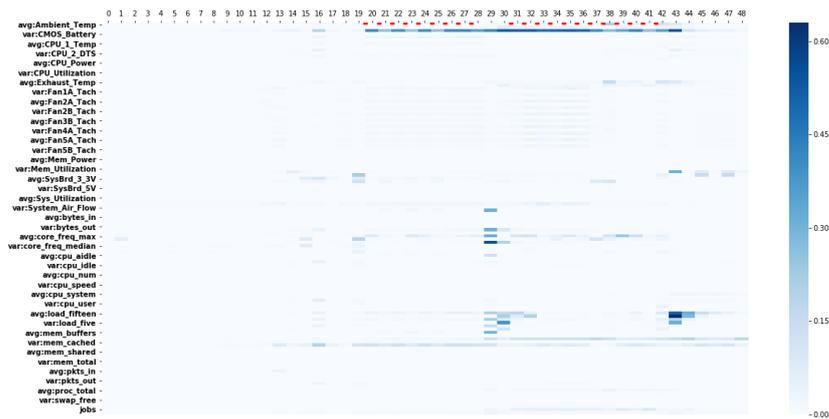


Figura 3.37: Heatmap 136c04s04 16 Gennaio

avg:core_freq_median	0.555181
avg:CMOS_Battery	0.412705
avg:core_freq_max	0.321791
avg:core_freq_avg	0.310634
avg:boottime	0.305890
avg:mem_buffers	0.298060
avg:load_five	0.213461
avg:load_fifteen	0.206180
avg:load_one	0.159247
avg:cpu_idle	0.137290

Tabella 3.15: Errori delle 18:25 dell'anomalia del 16 Gennaio di r136c04s04

r136c03s01

Il nodo in questione presenta l'anomalia del 16 Gennaio dalle 10:20 alle 10:40. Si può vedere nel grafico 3.38 che si ha il classico picco di errore nei periodi precedenti all'anomalia, circa un'ora prima e come si era già osservato si ha un picco nei periodi successivi dopo il buco di dati verso le 19:10. Durante l'anomalia l'errore di ricostruzione non ha un calo drastico come in altri casi.

L'AE-supervised per poco non riesce a superare la soglia di 0.5 per poter segnalare l'anomalia, però è evidente che rilevi quest'ultima, anche se in ritardo.

Nell'heatmap 3.39 si può evidenziare come siano presenti errori sulle metriche riguardante all'areazione (primo cerchio rosso in alto del plot) e un errore che è presente sin dalle ore 9:00 su "avg:proc_total" il quale rimane anche successivamente ma con errore minore. Successivamente ho evidenziato il secondo spike presente nel plot dell'errore, il quale presenta errori di ricostruzione su più features.

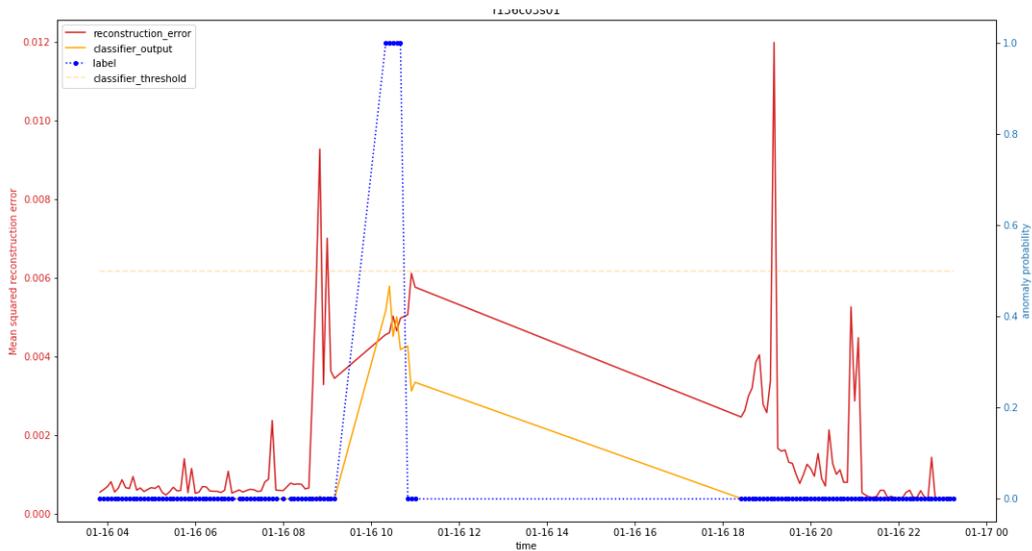


Figura 3.38: Comparazione AE-supervised AE-semisupervised r136c03s01 16 Gennaio

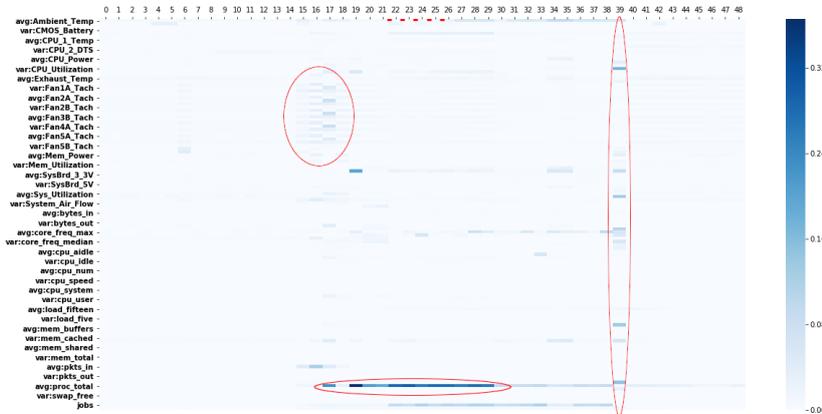


Figura 3.39: Heatmap r136c03s01 16 Gennaio

r130c12s03

Il nodo presenta l'anomalia del 16 Gennaio, dalle 10:20 alle 10:40, come il nodo analizzando precedentemente. Nel grafico 3.40 è presente il comune pico nell'errore prima che incominci l'anomalia, qui inoltre l'errore rimane molto alto anche durante il periodo anomalo. Successivamente al periodo anomalo l'errore rimane basso, indicando quindi un periodo di lavoro a regime.

AE-supervised identifica perfettamente l'anomalia, senza ritardo e senza anticipo superando 0.5 come output.

Nell'heatmap 3.41 si può notare come sia presente un errore persistente sulla metrica 'avg:mem_buffers' su tutto il periodo tranne in un punto in cui l'errore si sposta sulla metrica 'avg:boottime' la quale ho analizzato precedentemente e classificata come non necessaria.

È molto particolare il record 16 dell'heatmap che fa riferimento alle 9:00 nel quale è possibile identificare errori di ricostruzione sparsi su più features. Durante il periodo anomalo sono presenti multipli errori di ricostruzioni evidenti riportati in tabella 3.16

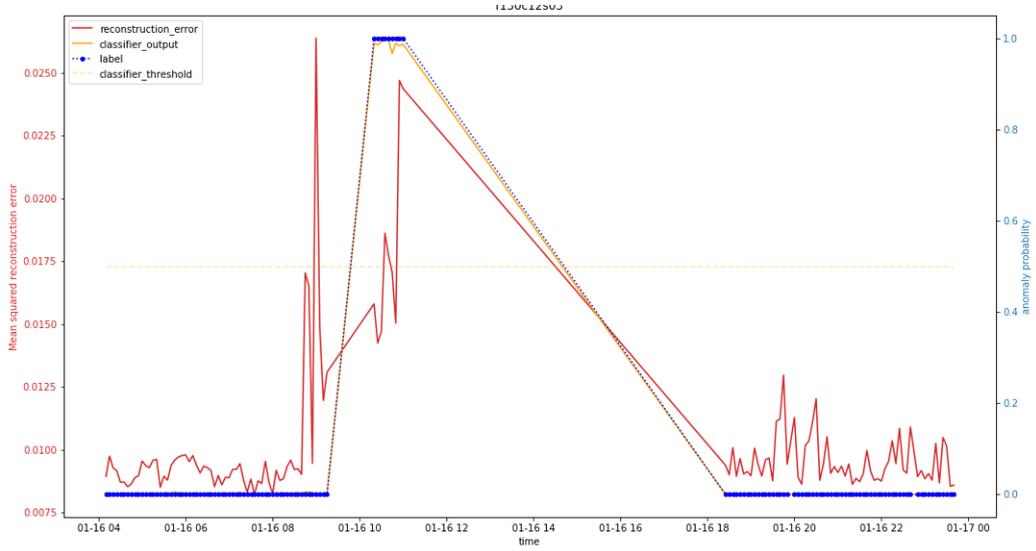


Figura 3.40: Comparazione AE-supervised AE-semisupervised r130c12s03 16 Gennaio

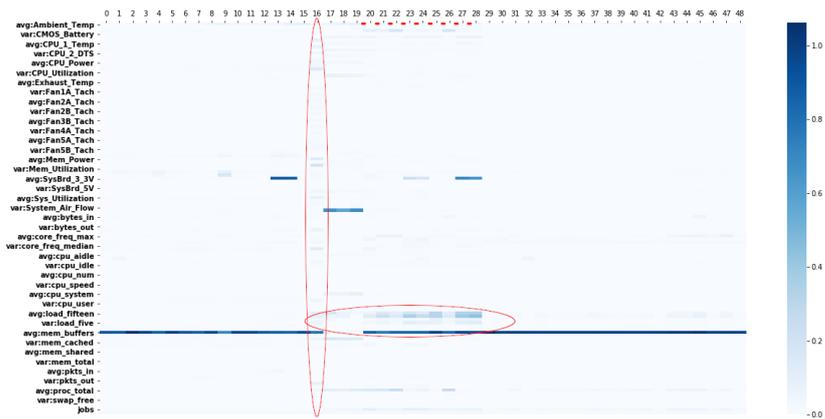


Figura 3.41: Heatmap r130c12s03 16 Gennaio

avg:mem_buffers	0.862230
var:load_fifteen	0.265595
avg:SysBrd_3_3V	0.202708
avg:load_fifteen	0.188495
avg:proc_total	0.099340
avg:CMOS_Battery	0.074614
var:load_five	0.071801
jobs	0.067724
avg:load_five	0.038979
avg:CPU_1_DTS	0.037806

Tabella 3.16: Errore durante anomalia del 16 Gennaio di r130c12s03

3.2.2 Analisi di Maggio

In questa sezione riporterò i grafici e le analisi compiute sul dataset dei vari nodi durante il mese di Maggio 2020.

r129c12s02

Questo nodo presenta una singola anomalia il 27 Maggio dalle 15:20 alle 15:25. Nel dataset è presente un buco nei dati dalle fine dell'anomalia (15:25) fino alle 15:50.

Come si può vedere in 3.42 le variazioni dell'errore sono tutti attorno all'anomalia, sia nel periodo precedente che quello successivo.

Osservando il grafico si può quindi dedurre che AE-semisupervised identifi chi il periodo anomalo in anticipo e rileva anche anomalie nelle features successivamente al periodo anomalo, probabilmente perché il nodo sta tornando a regime. AE-supervised invece rileva l'anomalia solamente in un record (di due) senza però superare la soglia di 0.5, fallendo quindi nel restituire la classe corretta in output.

Nell'heatmap 3.43 si può notare che gli errori siano tutti proprio attorno all'anomalia, gli errori in questione sono riguardati metriche delle CPU e del carico del nodo.

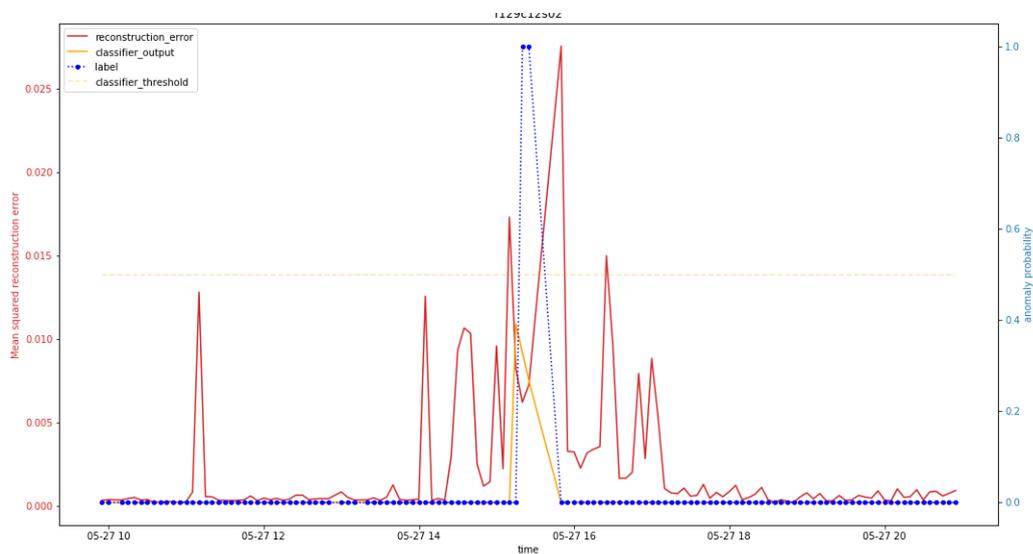


Figura 3.42: Comparazione AE-supervised AE-semisupervised r129c12s02 27 Maggio

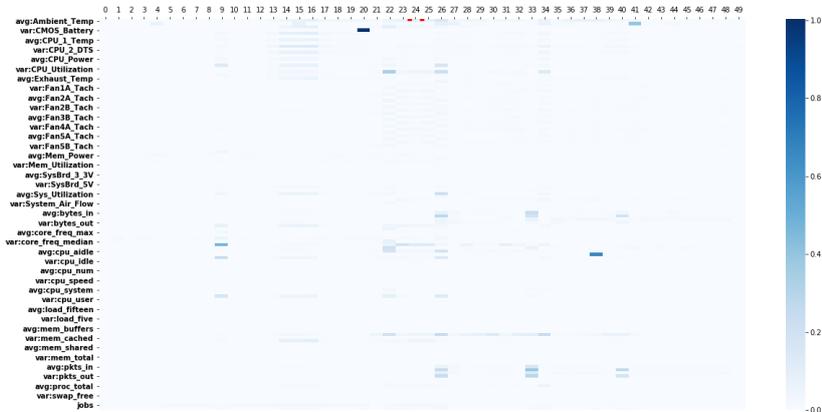


Figura 3.43: Heatmap r129c12s02 27 Maggio

r136c03s02

Il nodo ha una singola anomalia durante il periodo di Maggio, il 27 ma quest'anomalia è più longeva di quella descritta precedentemente poiché va dalle 15:20 alle 16:00, con un buco di dati dalle 15:25 alle 15:50 (esattamente come prima) ma in questo caso dalle 15:55 alle 16:00 è segnata anomalia da Nagios.

Nel grafico 3.44 si può osservare che l'errore precedentemente al picco vicino all'anomalia ha un comportamento oscillatorio. Poco prima dell'anomalia si ha un picco nell'errore e successivamente si ha un calo e nei periodi successivi si ha un elevato errore di ricostruzione. Per quanto riguarda AE-supervised identifica correttamente l'anomalia durante il periodo anomalo, ma nel periodo successivo anch'essa ha delle variazioni sull'output, minori di 0.5 per cui l'output della classe è corretto.

Nell'heatmap 3.45 è presente un errore molto marcato su più record prima dell'anomalia e fa riferimento alla metrica "avg:SysBrd_12V". Nella parte evidenziata prima dell'anomalia gli errori fanno tutti riferimenti a metriche di areazione del nodo (Ventole, flusso d'aria). L'errore medio durante l'anomalia è riportato nella tabella 3.17. Mentre dopo l'anomalia gli errori sono concentrati nel carico del nodo e temperatura delle CPU.

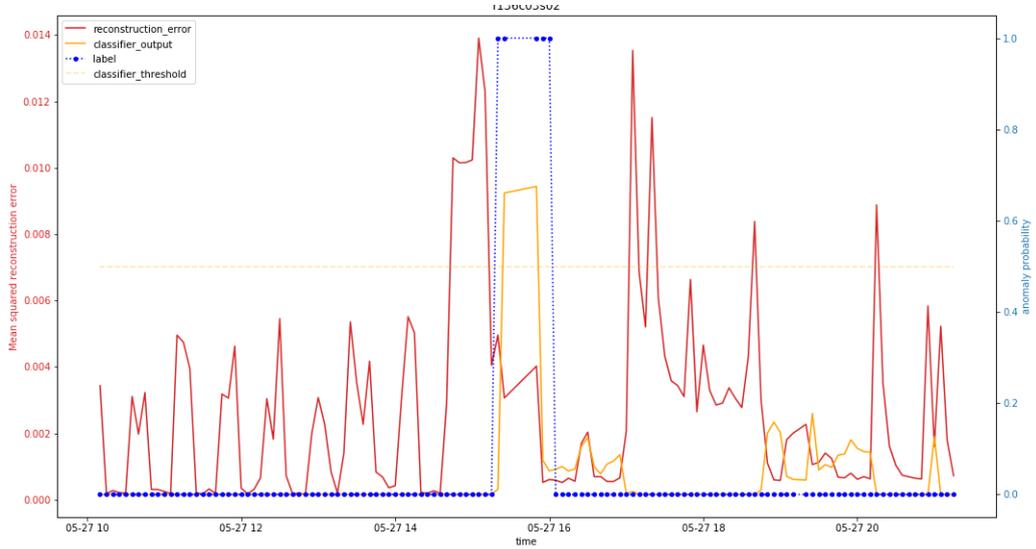


Figura 3.44: Comparazione AE-supervised AE-semisupervised r136c03s02 27 Maggio

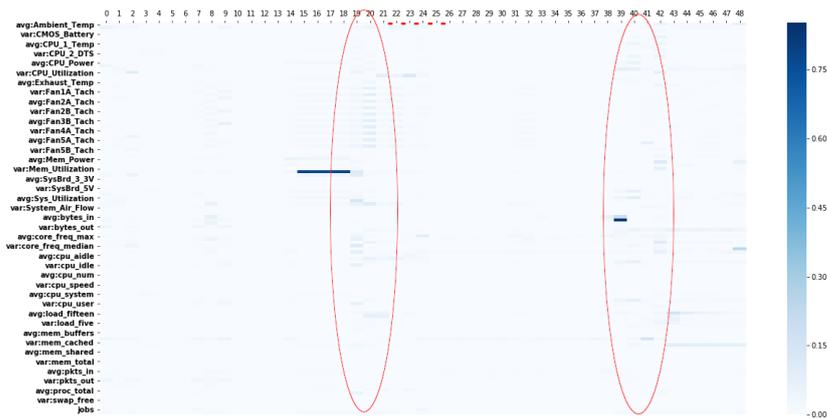


Figura 3.45: Heatmap r136c03s02 27 Maggio

avg:DC_Energy	0.040189
avg:core_freq_max	0.021167
var:Ambient_Temp	0.019170
var:cpu_idle	0.016921
avg:mem_cached	0.013116
avg:Fan3A_Tach	0.009389
avg:cpu_idle	0.009382
avg:Fan5B_Tach	0.009148
avg:Fan1B_Tach	0.009011
avg:Fan4B_Tach	0.008784

Tabella 3.17: Errore medio durante 27 Gennaio di r136c03s02

r141c09s02

Questo nodo non è stato analizzato durante l'analisi di Gennaio poiché non presentava anomalie durante il periodo. A Maggio invece presenta l'anomalia del 27 dalle 15:20 fino alle 15:25 (quindi due record) con il buco di dati dalle 15:25 alle 15:50.

Nel plot 3.46 l'errore prima dell'anomalia presenta un comportamento oscillatorio. Successivamente l'errore s'innalza nell'intorno dell'anomalia mentre nel periodo successivo si può vedere una chiara diminuzione del valore dell'errore nonostante preservi il comportamento oscillatorio. AE-supervised identifica correttamente l'anomalia con un output maggiore di 0.5.

Anche nell'heatmap 3.47 è possibile notare come prima dell'anomalia siano presenti molti errori sparsi su più features causando un incremento dell'errore dato che si utilizza il valore medio invece di un massimo. L'errore medio durante il periodo anomalo è riportato in tabella 3.18.

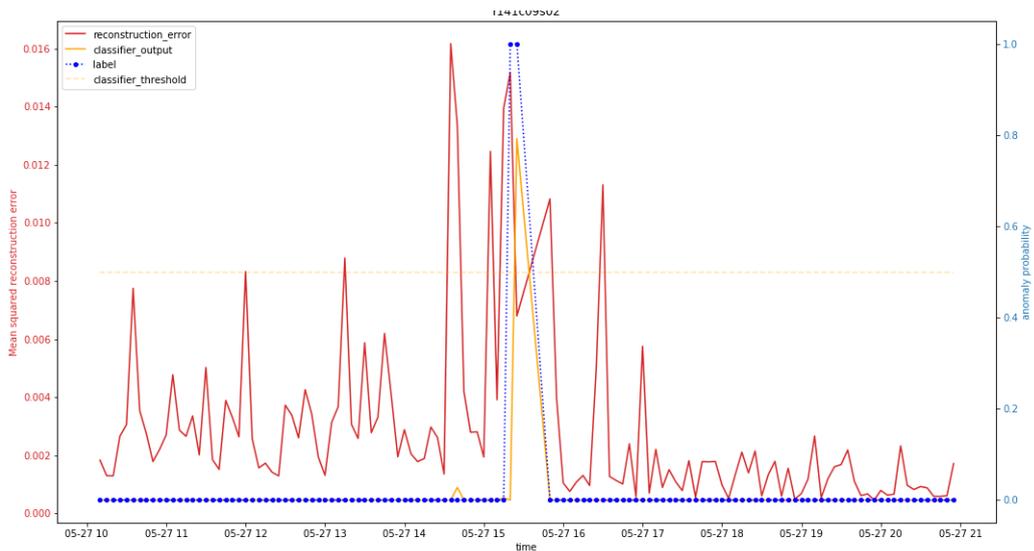


Figura 3.46: Comparazione AE-supervised AE-semisupervised r141c09s02 27 Maggio

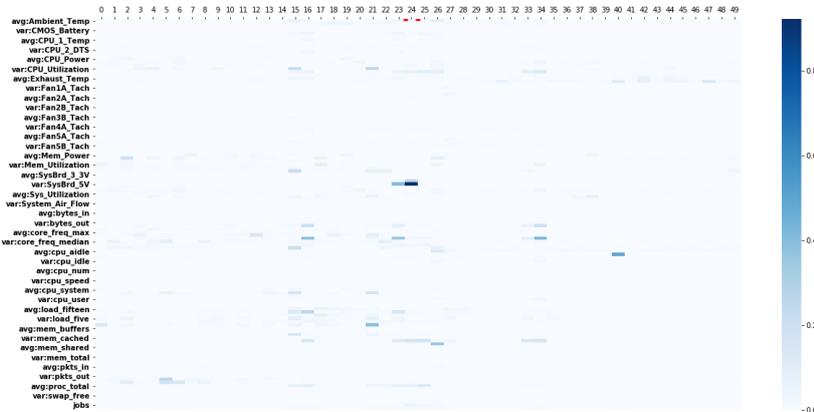


Figura 3.47: Heatmap r141c09s02 27 Maggio

var:SysBrd_5V	0.460933
avg:mem_free	0.159780
avg:SysBrd_5V	0.129746
avg:proc_total	0.108989
avg:DC_Energy	0.080235
jobs	0.043357
avg:core_freq_min	0.037191
avg:core_freq_median	0.034616
avg:Ambient_Temp	0.031025
avg:CPU_1_DTS	0.021769

Tabella 3.18: Errore medio anomalia del 27/05 di r141c09s02

r140c12s04

Il nodo presenta due anomalie, la prima è quella del 19 Maggio che va dalle 17:05 alle 17:15 senza buchi di dati e un'altra il 26 Maggio dalle 15:20 alle 15:25. Partirò analizzando quella del 19 Maggio.

Nel plot 3.48 si può già notare come sia presente un livello di errore costante e vicino all'anomalia (5 minuti prima, alle 17:00) si ha un picco notevole e qualche innalzamento spurio poco dopo la fine dell'anomalia. AE-supervised in quest'anomalia la identifica solamente in un record dei 3, in particolare in quello centrale con output maggiore di 0.5, non mostra segni di anticipo in quest'anomalia.

Nell'heatmap 3.49 sono chiaramente visibili gli errori permanenti di que-

var:cpu_wio	0.973963
var:proc_total	0.864982
var:load_one	0.665700
var:load_fifteen	0.461981
var:load_five	0.439650
var:core_freq_max	0.242197
avg:load_one	0.206632
var:core_freq_median	0.193450
avg:proc_total	0.179394
avg:cpu_wio	0.165576

Tabella 3.19: Errore pre-anomalia 19 Maggio r140c12s04

sto nodo che fanno riferimento alle metriche "avg:load_fifteen", "avg:load_five", "avg:proc_total". L'errore del record 22 (17:00) è visibile nella tabella 3.19 e come si può notare per la maggior parte sono errori che fanno riferimento alle varianze di quelle metriche. Mentre la media degli errori durante il periodo in tabella 3.20 fanno per la maggior parte riferimento alle medie delle stesse metriche.

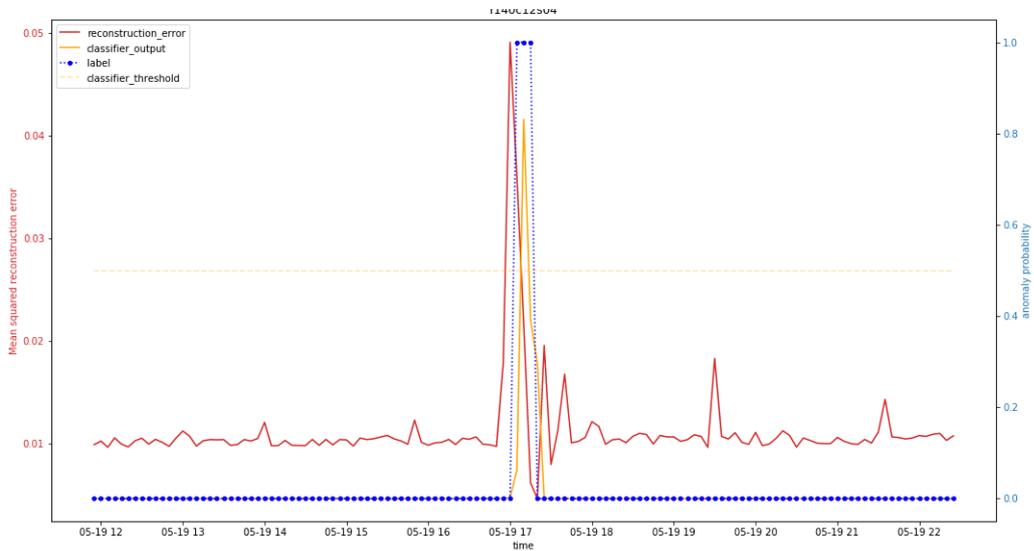


Figura 3.48: Comparazione AE-supervised AE-semisupervised r140c12s04 19 Maggio

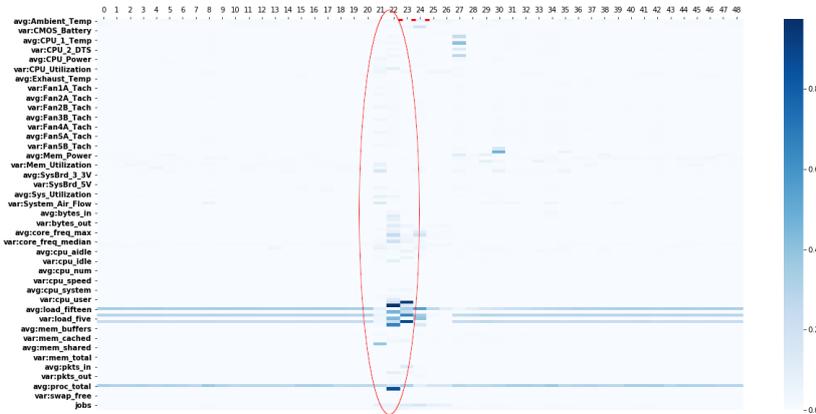


Figura 3.49: Heatmap r140c12s04 19 Maggio

avg:load_fifteen	0.378924
avg:load_five	0.344599
avg:cpu_wio	0.312741
avg:load_one	0.298910
avg:proc_total	0.196293
var:load_five	0.131656
jobs	0.118711
var:core_freq_max	0.078341
var:load_fifteen	0.071687
avg:CMOS_Battery	0.069271

Tabella 3.20: Errore durante l'anomalia 19 Maggio r140c12s04

Ora analizzo la seconda anomalia, comune anche ai nodi analizzati precedentemente in questo periodo.

In questo caso (Figura 3.50) non è presente un picco nell'immediate vicinanze dell'anomalia, ma alle 13:35 quindi 1h e 45 minuti prima dell'anomalia susseguito da altri innalzamenti ma di basso valore. In questo caso però si nota un comportamento nuovo, l'innalzamento dell'errore durante l'anomalia, fino ad'ora si era sempre osservato come durante l'anomalia l'errore calasse, qui si alza fino alle 15:50 quando ricala e assume un valore costante più alto dei valori presenti pre-anomalia.

Analizzando il heatmap 3.51 è possibile vedere come l'innalzamento dell'errore post-anomalia è dovuto all'aumento dell'errore di ricostruzione delle

feature che presentano un'errore di ricostruzione permanente viste anche nell'anomalia precedente. Il record 26 (associato alle 15:50) presenta errori elevati sulle metriche di areazione del nodo (ventole e flusso d'aria), gli errori evidenti segnalati nel record 3 riguardano il carico del nodo e infine durante il periodo anomalo si hanno errori su metriche di CPU quali temperatura e carico.

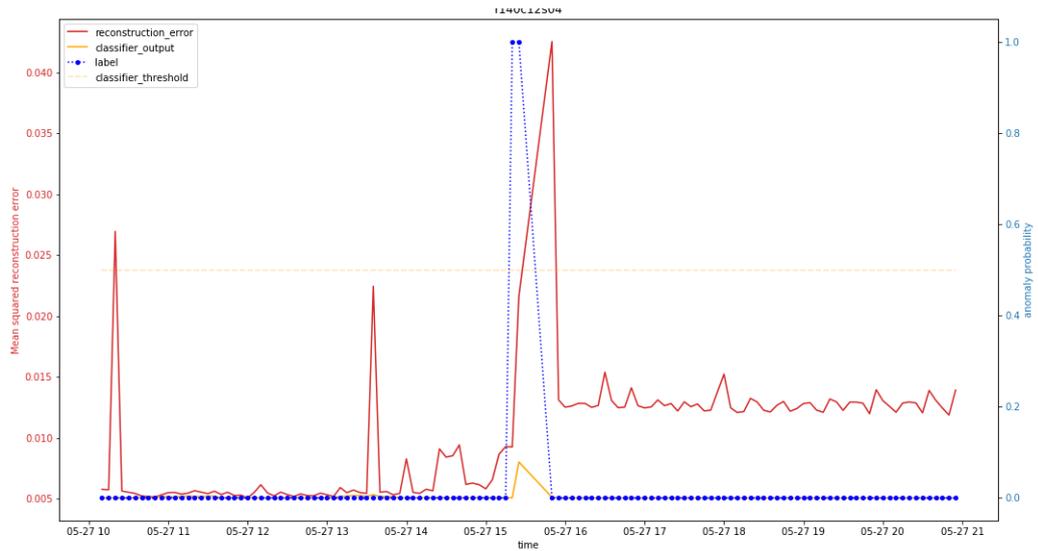


Figura 3.50: Comparazione AE-supervised AE-semisupervised r140c12s04 27 Maggio

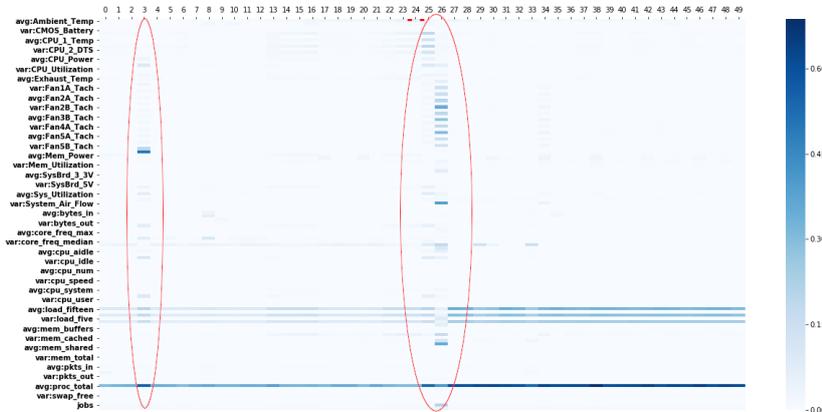


Figura 3.51: Heatmap r140c12s04 27 Maggio

r174c02s02

Questo nodo presenta l'anomalia del 27 Maggio dalle 15:20 alle 16:00 con il buco dalle 15:25 alle 15:50. In particolare da circa le 13:30 si alza l'errore fino alle 14:15 e successivamente risale poco prima dell'anomalia e si ha un calo durante il periodo anomalo. Nel periodo successivo all'anomalia si può vedere un comportamento oscillatorio con un trend di calo dell'errore.

Si può osservare anche come AE-supervised riesca a identificare l'anomalia con un anticipo di circa 10 minuti (2 record) con un output sopra la soglia di 0.5 e riesce anche ad avere un output alto durante il periodo anomalo stesso.

Analizzando la heatmap 3.53 si possono vedere errori evidenti su molte feature nei record a 0 a 8(13:30 - 14:10), e nei periodi successivi all'anomalia. In particolare:

- **Da 0 a 8:** Si hanno errori su "IO_Utilization", "CPU_Temp" e "CPU_DTS"
- **Da 15 a 19:** Errori su "SysBrd_12V" e "DC_Energy".
- **Durante l'anomalia:** "Exhaust_Temp", "proc_total", "DC_energy"
- **Dopo l'anomalia:** Errori sulla ventilazione del nodo

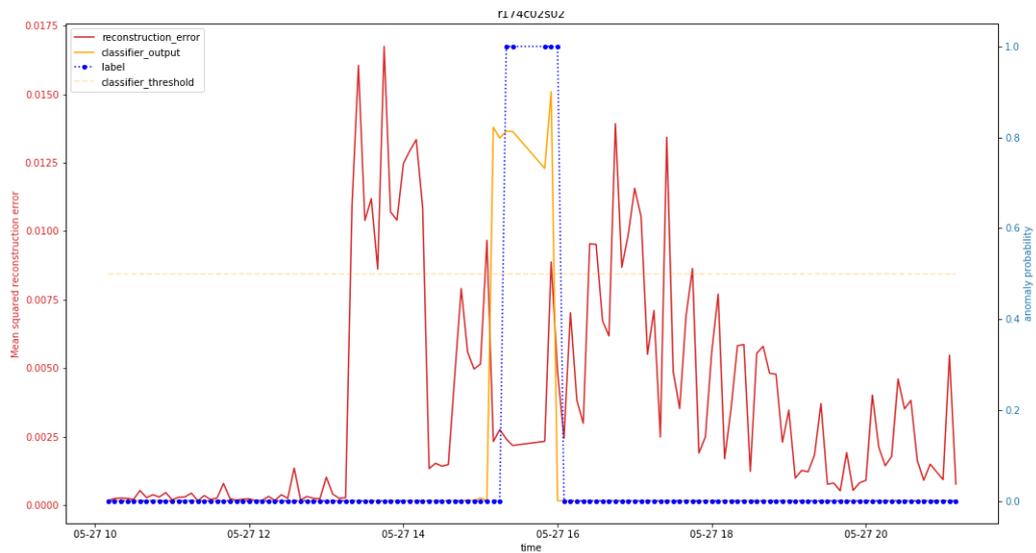


Figura 3.52: Comparazione AE-supervised AE-semisupervised r174c02s02 27 Maggio

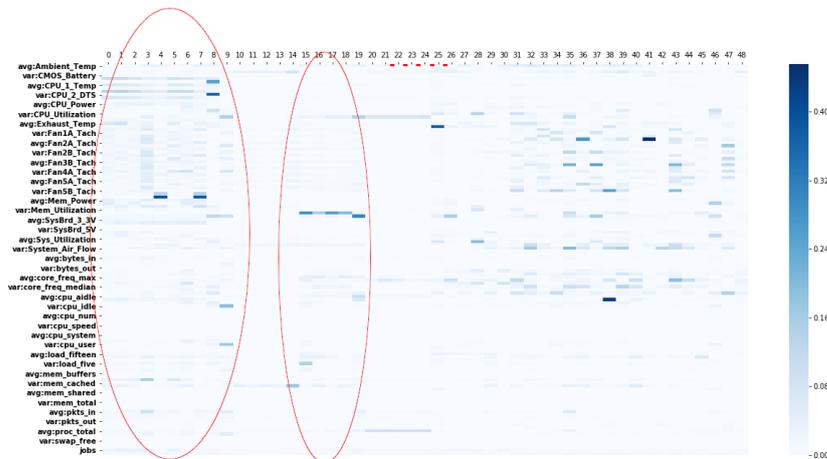


Figura 3.53: Heatmap r174c02s02 27 Maggio

r173c10s01

Questo nodo presenta l'anomalia del 27 Maggio dalle 15:20 alle 16:00 con un buco nei dati dalle 15:25 alle 15:50. Osservando l'heatmap 3.55 è possibile notare come gli errori inizino a essere più marcati nel periodo successivo all'anomalia. In particolare si può notare il record 32 (16:30) che mostra errori su metriche di areazione del nodo, mentre durante il periodo anomalo

gli errori sono mostrati in tabella 3.21 e come si può notare l'errore più marcato è quello su "Ambient_Temp".

Nel plot 3.54 si nota come non siano presenti picchi nell'errore nel periodo immediatamente precedente all'anomalia (come si è visto in molti casi). Qui AE-supervised inizia ad avere variazioni nel suo output circa alle 13 sebbene con un output molto basso ma da tenere in considerazione dato l'evidente sbilanciamento del dataset. AE-supervised riesce a identificare solamente un record anomalo e ha anche output rilevanti anche nelle fasi successive all'anomalia.

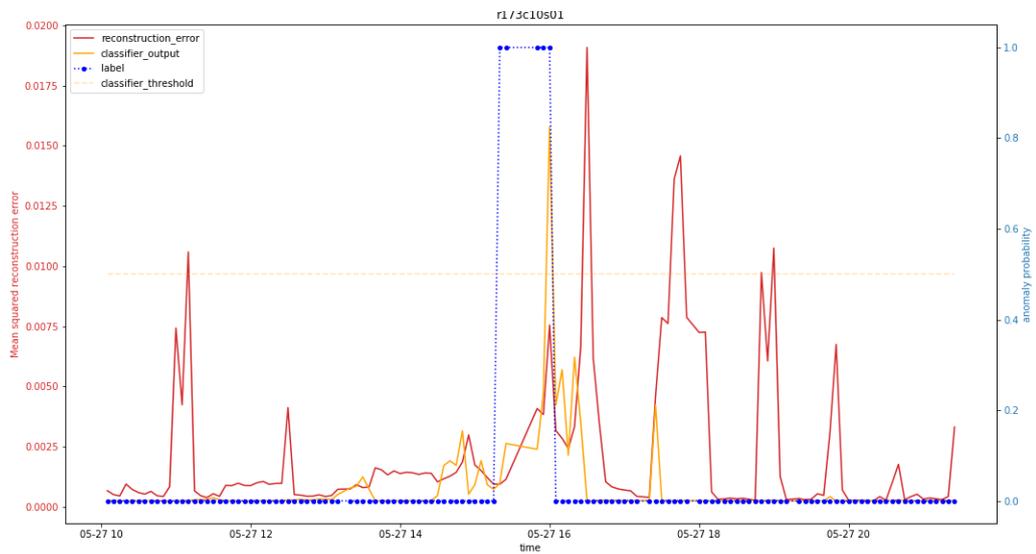


Figura 3.54: Comparazione AE-supervised AE-semisupervised r173c10s01 27 Maggio

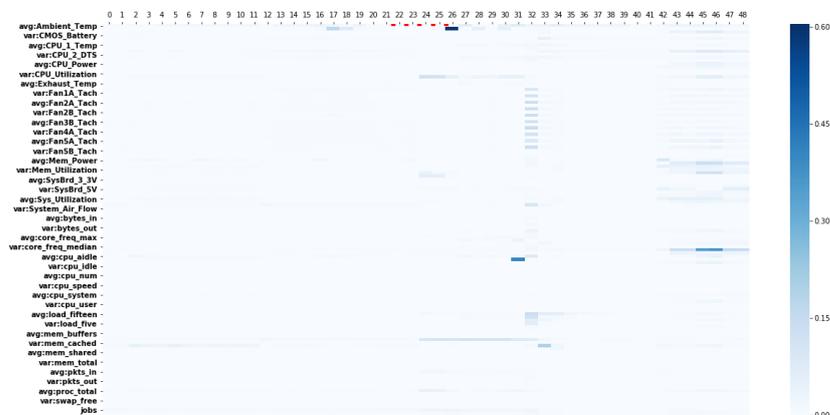


Figura 3.55: Heatmap r173c10s01 27 Maggio

var:Ambient_Temp	0.124472
avg:mem_cached	0.065312
avg:DC_Energy	0.057988
var:SysBrd_12V	0.023667
avg:proc_total	0.023437
jobs	0.022834
avg:SysBrd_12V	0.012812
avg:pkts_in	0.009510
var:SysBrd_5V	0.005727
avg:mem_free	0.005176

Tabella 3.21: Errore durante il 27 Maggio di r173c10s01

r130c12s04

Questo nodo non è presente nell'analisi di Gennaio poiché non mostrava anomalie nel periodo. In Maggio invece presenta due anomalie, una il 27 Maggio con un singolo record alle 15:20 e una il 25 Maggio dalle 20:50 alle 6:00 del 26.

Partendo dall'anomalia più comune nel dataset (27 Maggio) si nota che è presente il classico buco di dati dalle 15:25 alle 15:50 comprese.

Analizzando il plot dell'errore e dell'output di AE-supervised si nota come dalle 14:30 l'errore medio si alzi, allo stesso istante anche l'output di AE-supervised inizia a cambiare, arrivando anche a superare la soglia di 0.5 poco prima dell'anomalia.

L'errore dalle 14:30 è in crescita fino al record successivo all'anomalia dove assume il valore massimo, durante l'anomalia AE-supervised identifica correttamente l'anomalia con un output di 0.5.

Nell'heatmap 3.57 ho evidenziato 3 sezioni che riportano errori che incominciano alle 14:30 circa e continuano fino alle 16:15 circa. Questi errori riguardano le seguenti metriche (dall'alto verso il basso): metriche di ventilazione (ventole in particolare), metriche riguardano la frequenza della cpu, carico del nodo (load_five e load_fifteen).

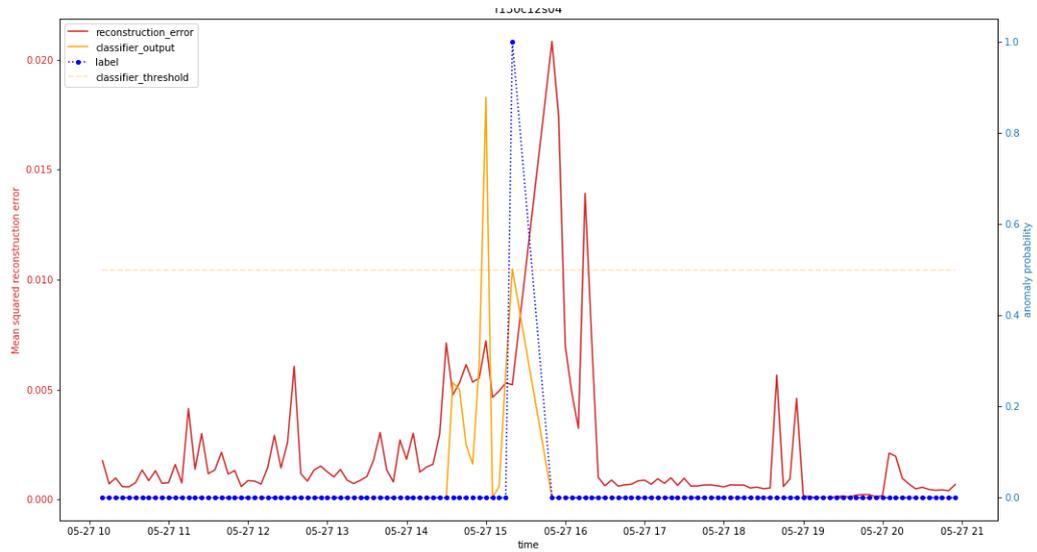


Figura 3.56: Comparazione AE-supervised AE-semisupervised r130c12s04 27 Maggio

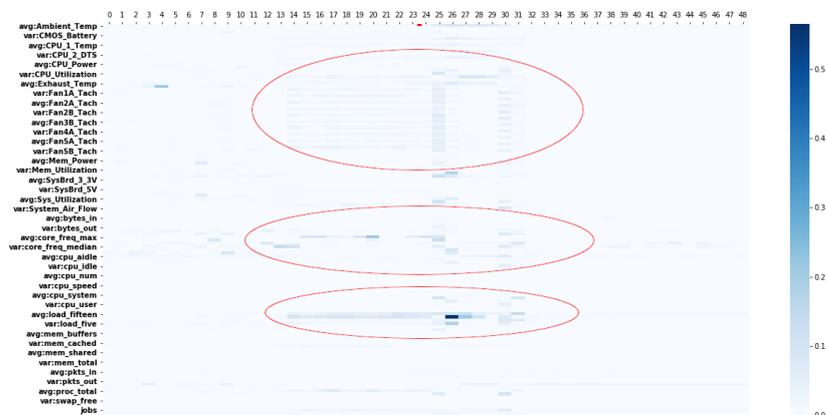


Figura 3.57: Heatmap r130c12s04 27 Maggio

La seconda anomalia è più lunga di quelle riportate precedentemente quindi andrò a riportare le heatmap delle parti più salienti ovvero inizio e fine. Ricordo che l'anomalia dura dalle 20:50 del 25 Maggio alle 6:00 del 26 Maggio.

Analizzando il plot che raffigura l'inizio dell'anomalia (3.58) si può notare come alle 18:00 del 25 l'errore di ricostruzione medio ha un innalzamento significativo fino alle 20; data la durata il comportamento non oscillatorio potrebbe essere associato a un primo sintomo di anomalia. Successivamente l'errore ha un calo fino all'inizio dell'anomalia dove l'errore di ricostruzione ha un picco notevole nel record precedente all'inizio, inoltre l'errore rimane elevato durante l'anomalia, sebbene con un comportamento oscillatorio.

Per quanto riguarda AE-supervised, dall'inizio dell'anomalia ha una sostanziale variazione dell'output, passa da 0 a 1 in un singolo record. Durante l'anomalia in due situazione ha due cali sotto la soglia di 0.5 ma i punti classificati correttamente sono la grande maggioranza in questa fase iniziale.

Guardando anche l'heatmap 3.59 che rappresenta il periodo (18:45 - 22:50) possiamo associare il picco dovuto alle ore precedente all'anomalia all'elevato errore di ricostruzione sulla metrica "jobs". È evidente anche come gli errori inizino a essere più evidenti durante l'anomalia. In particolare questi errori sono molto marcati su metriche riguardanti la cpu (Tabella 3.22).

La parte centrale dell'anomalia presenta errori di ricostruzione sempre sulle stesse features, riporterò quindi solo la tabella dell'errore (Tabella 3.23).

Nell'heatmap riguardante la parte finale dell'anomalia (Figura 3.60) ho evidenziato le metriche che smettono di aver errori di ricostruzione dopo la

avg:core_freq_max	0.143056
avg:cpu_idle	0.094518
var:SysBrd_12V	0.041412
avg:CPU_1_Temp	0.040675
var:core_freq_max	0.035985
avg:CPU_Power	0.030351
avg:Exhaust_Temp	0.025675
avg:Sys_Power	0.024807
avg:load_fifteen	0.022988
var:load_fifteen	0.021961

Tabella 3.22: Errore medio durante l’inizio dell’anomalia del 25 di r130c12s04

fine dell’anomalia quali (avg e var di load_fifteen e avg:proc_total). Inoltre ho anche evidenziato che successivamente all’anomalia la metrica ”avg:CPU_Utilization” inizia a presentare un errore di ricostruzione più elevato (cerchio rosso in alto a sinistra).

Nella parte finale dell’anomalia nel plot (Figura 3.61) si può vedere come l’errore cali solamente qualche ora dopo la fine dell’anomalia e come AE-supervised cali l’output drasticamente alla fine dell’anomalia.

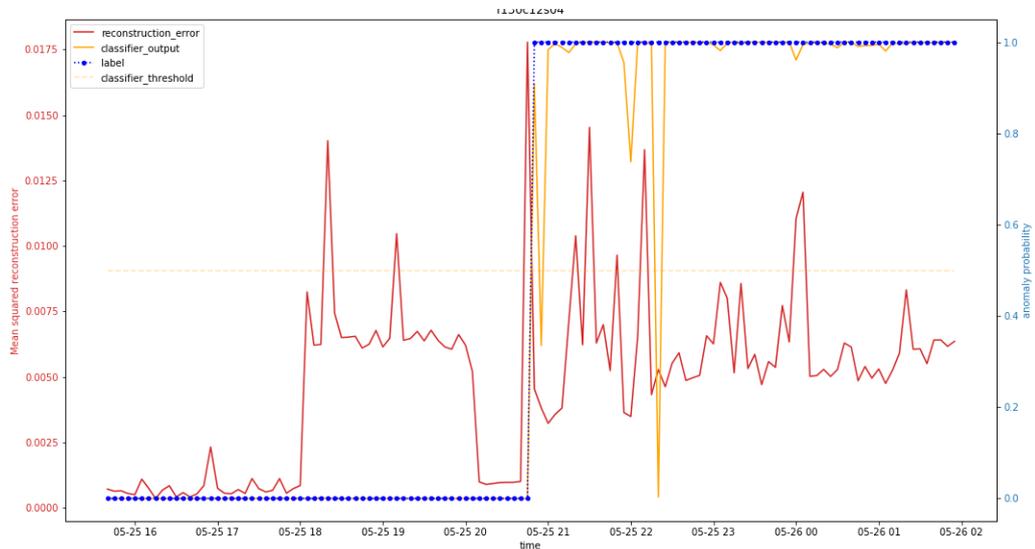


Figura 3.58: Comparazione AE-supervised AE-semisupervised r130c12s04 25 Maggio - Inizio Anomalia

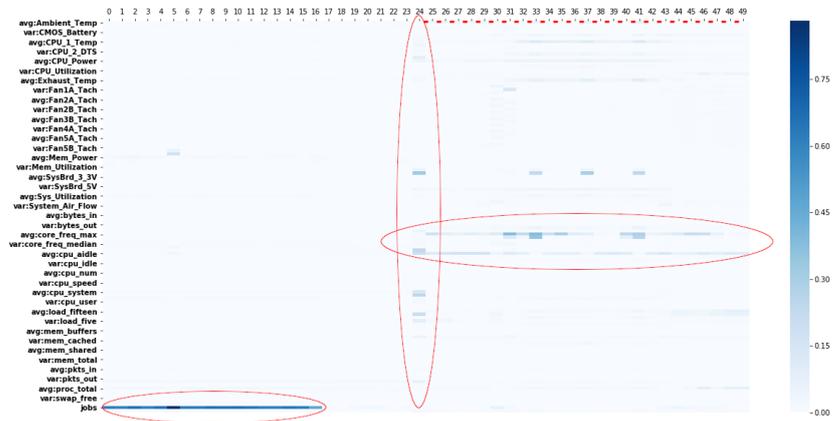


Figura 3.59: Heatmap r130c12s04 25 Maggio - Inizio Anomalia

avg:core_freq_max	0.148842
avg:cpu_idle	0.072497
avg:load_fifteen	0.058359
var:load_fifteen	0.051636
avg:proc_total	0.047446
avg:DC_Energy	0.025408
var:core_freq_max	0.024519
var:cpu_idle	0.022297
avg:Exhaust_Temp	0.016216
jobs	0.015893

Tabella 3.23: Errore medio durante l'anomalia del 25 di r130c12s04

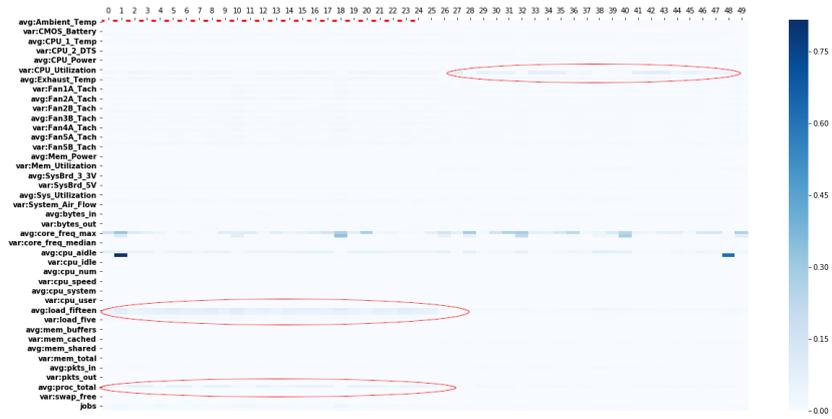


Figura 3.60: Heatmap r130c12s04 25 Maggio - Fine Anomalia

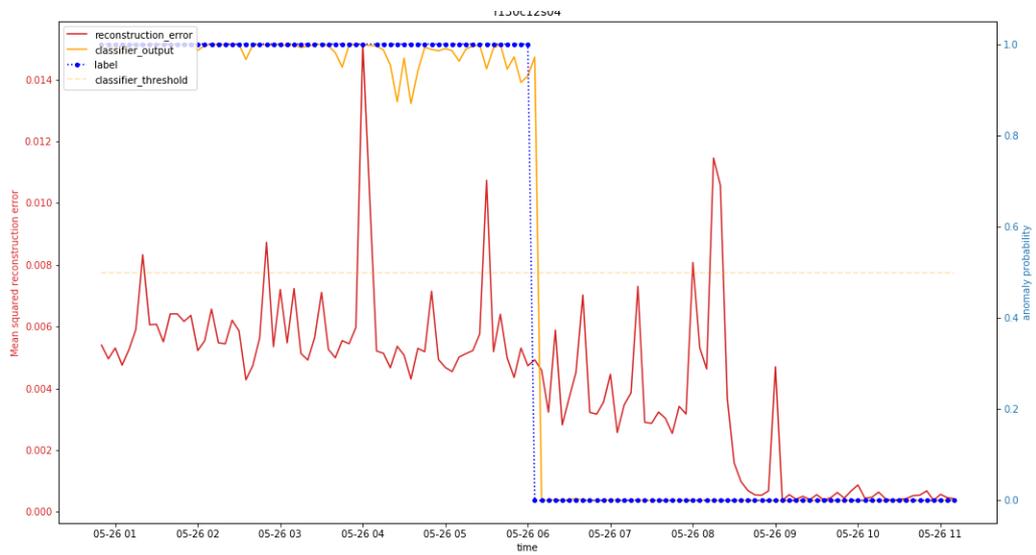


Figura 3.61: Comparazione AE-supervised AE-semisupervised r130c12s04 25 Maggio - Fine Anomalia

r162c05s02

r183c05s01

Il nodo in questione presenta un'anomalia il 27 Maggio dalle 15:20 alle 16:00 con un buco di dati dalle 15:25 alle 15:50.

In questo caso (Figura 3.62) non è presente un picco significativo nell'errore prima del periodo anomalo, ma si può notare un picco nell'output di AE-supervised però inferiore alla soglia di 0.5. Durante l'anomalia si ha un incremento dell'errore di ricostruzione e AE-supervised ha un incremento nel suo output ma non riesce a raggiungere una soglia.

Nei periodi successivi all'anomalia l'output di AE-supervised non torna a zero e nemmeno l'errore di ricostruzione medio che ha un picco intorno alle ore 17:45, si può notare inoltre che anche dopo le ore 20 AE-supervised ha un output oscillatorio.

L'heatmap (Figura 3.63) presenta errori di ricostruzione dopo l'inizio dell'anomalia. In particolare sulle features: "avg:DC_Energy", "avg:proc_total". Nel heatmap ho evidenziato il record appartenente al picco delle 17:45 che ha come features con errore di ricostruzione più alto le seguenti: "var e avg mem_free", "avg:mem_cached" e "avg:load_fifteen".

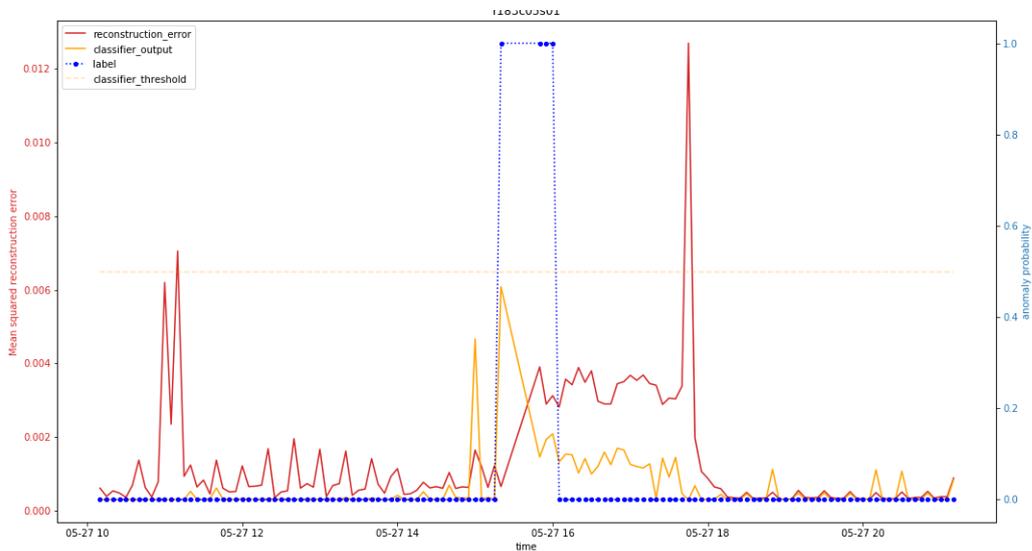


Figura 3.62: Comparazione AE-supervised AE-semisupervised r183c05s01 27 Maggio



Figura 3.63: Heatmap r183c05s01 27 Maggio

r129c12s03

Anche questo nodo presenta solo l’anomalia del 27 Maggio dalle 15:20 fino alle 16:00 con il buco di dati dalle 15:25 alle 15:50.

L’errore prima dell’anomalia è sempre molto basso, inizia ad alzarsi poco prima del periodo anomalo (Figura 3.64). AE-supervised identifica correttamente l’anomalia.

Si hanno picchi di errore di ricostruzione negli istanti successivi alla fine dell’anomalia che hanno un trend di decrescita, al fine di capire su che features sono presenti questi errori analizzo anche la heatmap.

Nell’heatmap (Figura 3.65) è possibile notare come fino a poco prima del periodo anomalo non siano presenti errori di ricostruzione e questi incomincino dal record 22 che fa riferimento alle 15:10. Gli errori di ricostruzione durante il periodo anomalo fanno riferimento a metriche di ventilazione del nodo e anche all’alimentazione con "DC_Energy", errore che è presente anche nei record successivi.

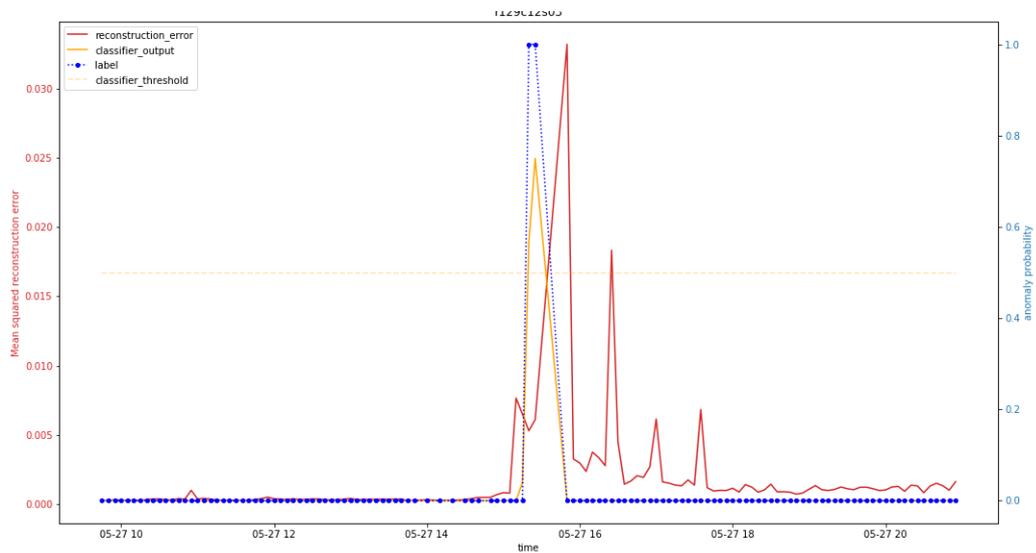


Figura 3.64: Comparazione AE-supervised AE-semisupervised r129c12s03 27 Maggio

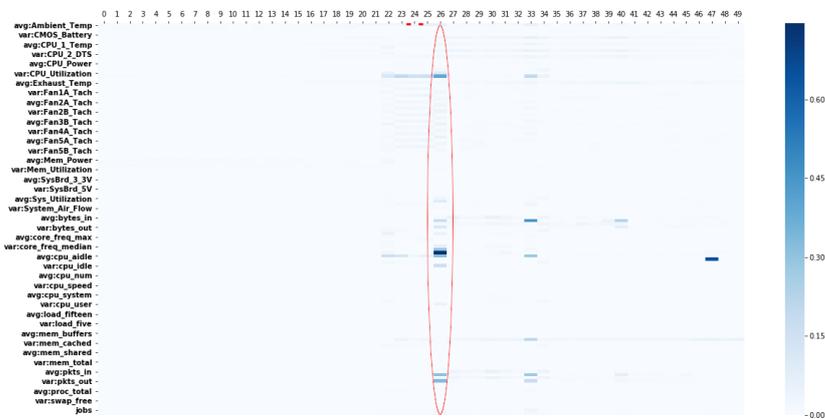


Figura 3.65: Heatmap r129c12s03 27 Maggio

r141c09s01

Il nodo presenta l'anomalia del 27 Maggio dalle 15:20 alle 15:25 col buco di dati nei periodi successivi (15:25-15:50).

Si può notare un picco spurio molto lontano dall'anomalia a livello temporale, data l'isolamento di tale picco non verrà considerato in termini di "anticipo" dell'anomalia.

Nei dintorni dell'anomalia invece ci sono 2 picchi nell'errore di ricostruzione (prima e dopo l'anomalia), guardando bene si ha un trend d'innalzamento dell'errore nei periodi precedenti fino ad avere il picco in questione.

Nell'heatmap 3.67 si può vedere che gli errori siano concentrati nell'intorno dell'anomalia. In particolare già dal record 21 (15:05) si evidenzia già un record con errori di ricostruzioni multipli con maggior valore su: "var:mem_free" ed errori più lievi su metriche di areazione del nodo. Mentre nel periodo successivo all'anomalia ci sono errori riguardo alle metriche della CPU.

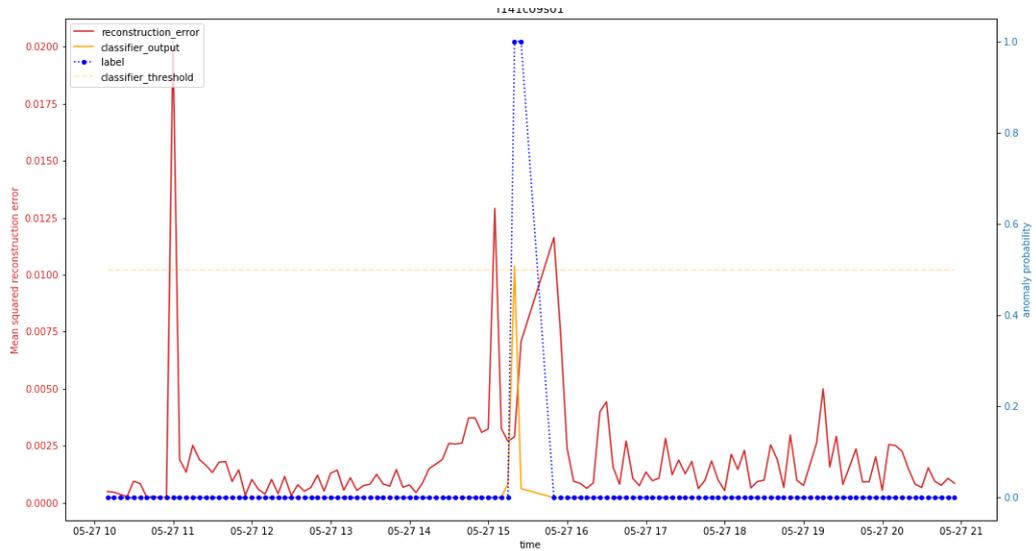


Figura 3.66: Comparazione AE-supervised AE-semisupervised r141c09s01 27 Maggio

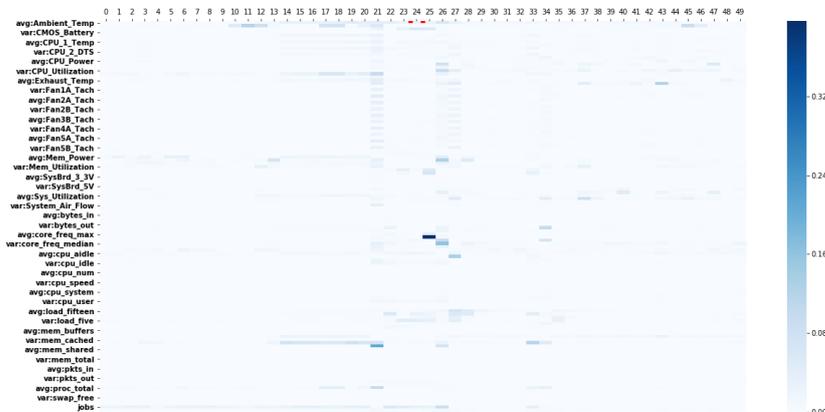


Figura 3.67: Heatmap r141c09s01 27 Maggio

r135c09s03

Il nodo presenta l’anomalia del 27 Maggio dalle 15:20 alle 15:25 con buco di dati dalle 15:25 alle 15:50.

Dal plot (Figura 3.68) si può notare come dalle 14 sia presente un primo picco dell’errore medio di ricostruzione che nei minuti successivi incrementa, anche nei periodi successivi all’anomalia nel quale sono presenti altri picchi fino ad arrivare a una situazione di regime. È evidente che gli errori siano concentrati nel periodo anomalo.

Per quanto riguarda la performance di AE-supervised si può notare che né durante il periodo anomalo né nei periodi precedenti e successivi ha variazioni nell’output, quindi non identifica l’anomalia in nessun modo.

Nell’heatmap (Figura 3.69) si nota come prima dell’anomalia siano già presenti errori di ricostruzioni sparsi su più feature, in particolare dal record 12 che fa riferimento alle 14:20, da 14 invece si ha un errore che continuerà fino a un record successivo all’anomalia su "avg:CMOS_Battery" che è l’errore più marcato nell’intero plot.

Durante l’anomalia ho evidenziato 2 errori di magnitudine minore oltre a quello già indicato precedentemente, questo errore è sulle metriche di "avg:proc_total" e "avg:cpu_idle".

Inoltre nel record 26 che rappresenta il primo record "normale" successivo all’anomalia che si riferisce alle 15:50 presenta errori di ricostruzioni multipli, in particolare sempre su "avg:CMOS_Battery" ma anche su metriche di carico di nodo come ad esempio "avg:load_fifteen", "avg:load_five" e

anche metriche relative all'alimentazione del nodo come "avg:DC_Energy" e "avg:SysBrd_12V". Data la natura di questo record e la label assegnata si potrebbe trattare di un record che rappresenta la transizione da anomalia a comportamento a regime.

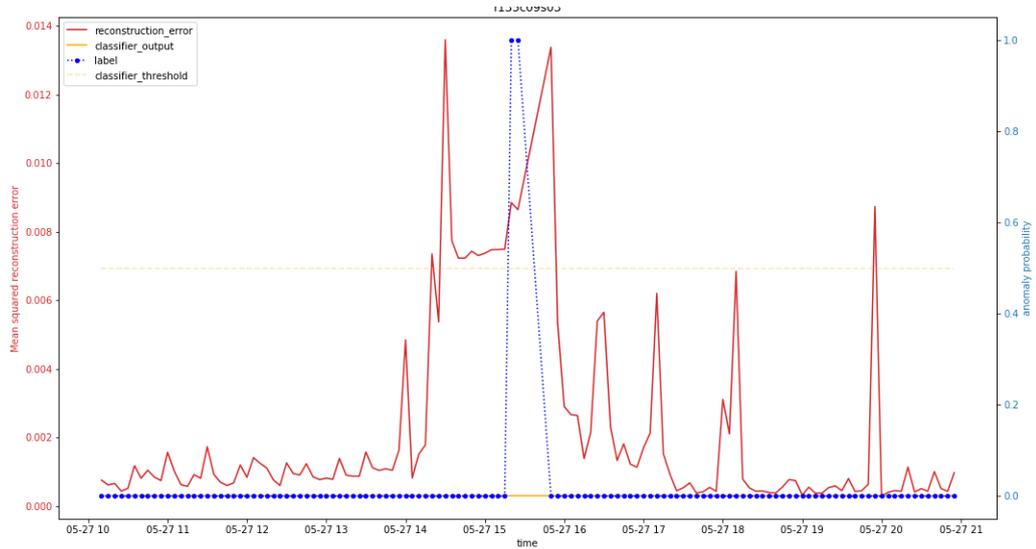


Figura 3.68: Comparazione AE-supervised AE-semisupervised r135c09s03 27 Maggio

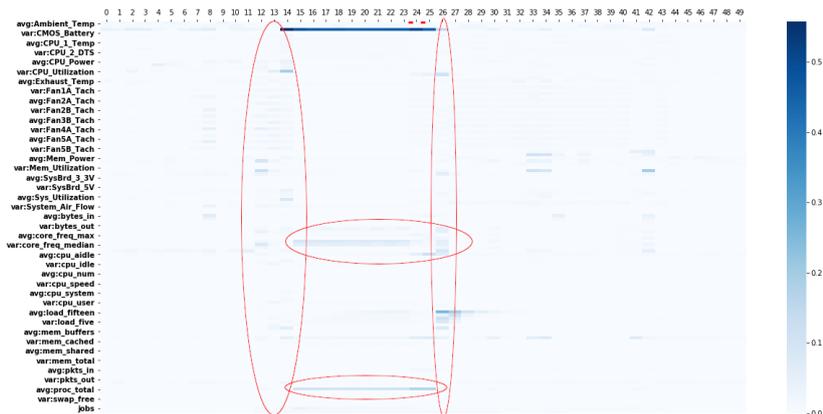


Figura 3.69: Heatmap r135c09s03 27 Maggio

r136c04s04

Il nodo presenta l'anomalia del 27 Maggio con soli due record anomali alle 15:20 e alle 15:25, è presente il buco di dati dalle 15:25 alle 15:50.

Sebbene possa sembrare la solita anomalia del 27 Maggio, qui si ha un comportamento molto particolare di AE-supervised il quale presenta un picco nell'output oltre 0.5 circa alle 13:30 e successivi innalzamenti dell'output poco dopo le 14:30 qui però senza superare la soglia. Totalmente differente l'andamento dell'errore di ricostruzione medio poiché non presenta spike evidenti prima dell'anomalia, ma solamente alle 15:50 e qualcuno spurio successivamente all'anomalia.

Osservando l'heatmap relativa a quest'anomalia (Figura 3.71) si nota come nei periodi precedenti all'anomalia, in particolare dal record 0 al record 16 (dalle 13:20 alle 14:40) sia presente un errore di ricostruzione "continuo" sulla metrica "var:SysBrd_12V".

Durante l'anomalia l'unico errore di ricostruzione che emerge con un valore nettamente superiore agli altri (0.021 mentre gli altri sono minori di 0.007) è la metrica "avg:SysBrd_12V", una metrica quindi strettamente collegata all'errore di ricostruzione rilevato nei periodi precedenti.

Il record 26 (15:50) presenta errori di ricostruzioni multipli, anche in questo caso è il primo record sano successivo all'anomalia, gli errori sono riportati in tabella 3.24.

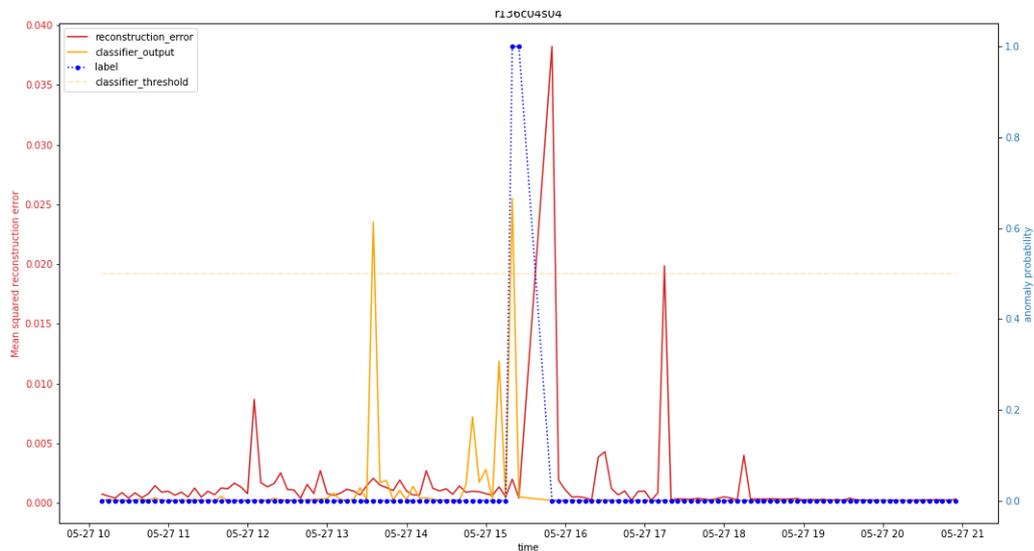


Figura 3.70: Comparazione AE-supervised AE-semisupervised r136c04s04 27 Maggio

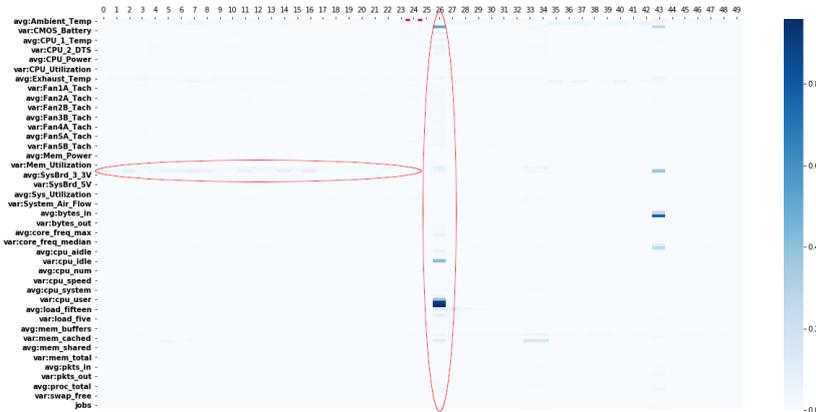


Figura 3.71: Heatmap r136c04s04 27 Maggio

var:cpu_wio	0.958633
avg:cpu_wio	0.905457
avg:CMOS_Battery	0.507496
var:cpu_idle	0.405620
var:cpu_user	0.377234
avg:load_fifteen	0.114742
avg:load_five	0.105999
avg:mem_free	0.103847
avg:SysBrd_12V	0.085305
avg:CPU_1_DTS	0.069179

Tabella 3.24: Errore 15:50 27 Maggio di r136c04s04

r183c05s02

Il nodo presenta due anomalie, la classica del 27 Maggio dalle 15:05 alle 16:00 col buco di dati dalle 15:25 alle 15:50 e un'altra anomalia lunga il 29 Maggio dalle 00:35 al 01/06 poiché il dataset finisce col primo giorno di Giugno.

Parto analizzando l'anomalia del 27 Maggio.

La prima parte del plot (Figura 3.72) poco dopo le 10 c'è uno spike dell'errore e anche dell'output di AE-supervised, con una successiva decrescita del trend dell'errore di ricostruzione. Dalle 14:00 l'errore di ricostruzione incrementa fino con comportamento oscillatorio fino alle 18:00.

Si può notare che AE-supervised inizia ad avere variazioni dell'output

avg:DC_Energy	0.500614
avg:cpu_idle	0.216318
avg:mem_free	0.174801
avg:proc_total	0.104412
avg:pkts_in	0.088333
avg:Ambient_Temp	0.034298
avg:core_freq_max	0.031310
avg:CMOS_Battery	0.028938
avg:core_freq_median	0.023285
avg:core_freq_avg	0.021782

Tabella 3.25: Errore medio durante l’anomalia del 27 Maggio di r183c05s02

sotto la soglia proprio dopo la fine dell’anomalia.

È molto interessante osservare anche la heatmap (Figura 3.73) dove sono visibili molti errori di ricostruzioni non particolarmente sparsi, ma su features specifiche riportate nella tabella 3.25. Nell’heatmap ho evidenziato che durante l’anomalia termina l’errore di ricostruzione su ”avg:mem_free” e inizia un errore di ricostruzione su ”avg:proc_total”.

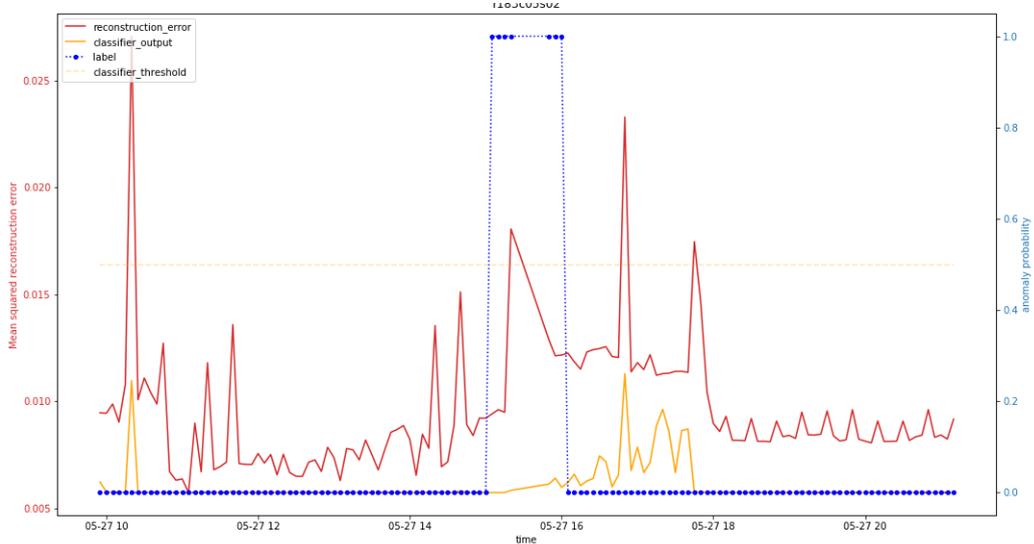


Figura 3.72: Comparazione AE-supervised AE-semisupervised r183c05s02 27 Maggio

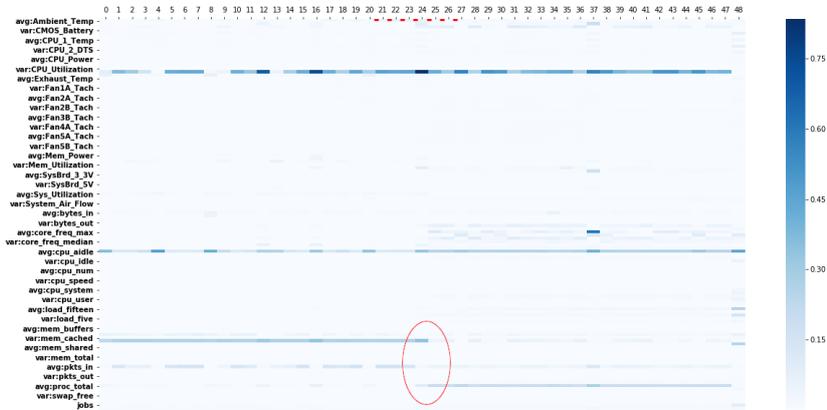


Figura 3.73: Heatmap r183c05s02 27 Maggio

In questo caso particolare è molto più significativa analizzare l'heatmap relativa all'inizio, poiché la parte centrale e la fine sono identiche all'inizio e per di più la fine non è completa a causa della finestra temporale ricoperta dal dataset.

Nell'heatmap (Figura 3.75 è evidente l'inizio dell'anomalia, appena l'anomalia inizia si creano errori di ricostruzione con valori molto alti su due metriche in particolare "avg:boottime" e "avg:mem_buffers". L'errore su "Boottime" come analizzato precedentemente è dovuto al reboot del nodo.

Nei periodi precedenti all'anomalia sono presenti altri due errori di ricostruzione ricorrenti fino all'inizio dell'anomalia, in particolare su "avg:cpu_idle" e "avg:mem_cached".

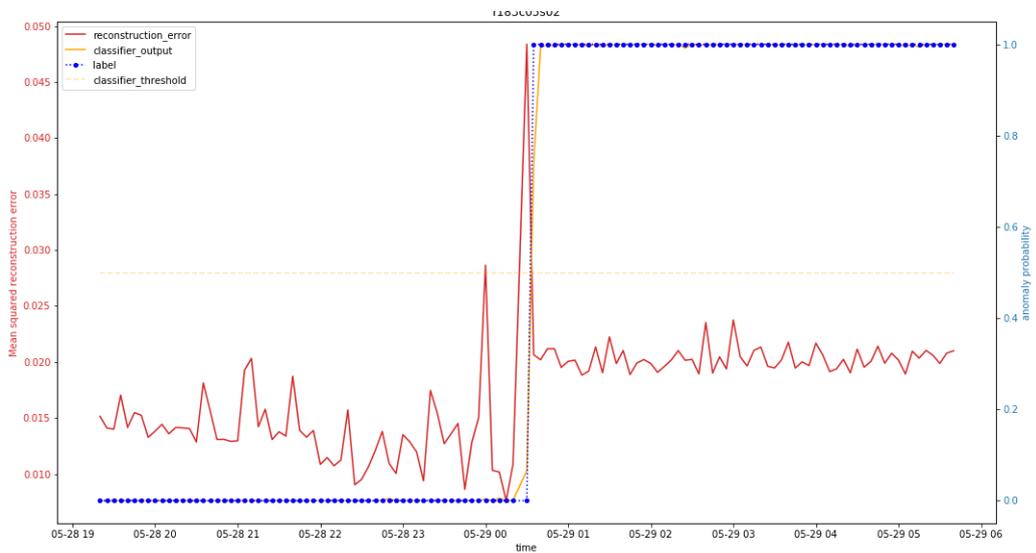


Figura 3.74: Comparazione AE-supervised AE-semisupervised r183c05s02 29 Maggio - Inizio

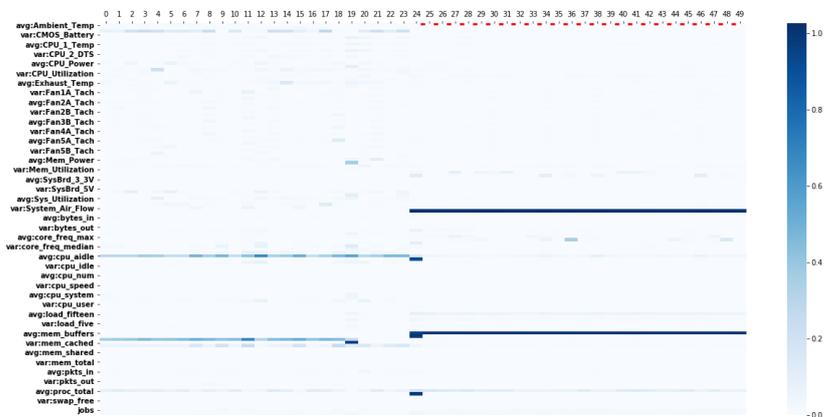


Figura 3.75: Heatmap r183c05s02 29 Maggio

r174c02s01

Il nodo presenta l'anomalia del 27 Maggio dalle 15:05 alle 16:00 con il buco di dati dalle 15:25 alle 15:50.

Come nelle analisi precedenti gli errori di ricostruzione sono concentrati intorno all'anomalia, ma coi valori più bassi proprio durante l'anomalia.

Per quanto riguarda AE-supervised identifica correttamente l'anomalia

avg:CMOS_Battery	0.563964
avg:pkts_out	0.436706
avg:bytes_out	0.400733
avg:bytes_in	0.330493
var:core_freq_min	0.316104
avg:DC_Energy	0.223476
var:core_freq_max	0.114515
avg:cpu_idle	0.105360
var:pkts_out	0.086184
var:core_freq_median	0.086178

Tabella 3.26: Errore 13:20-14:20 di r174c02s01 27 Maggio

con un output molto alto, con anche un falso positivo in un recor successivo all'anomalia.

Analizzando l'heatmap (Figura 3.77) è possibile notare come dal record a 0 al 12 (dalle 13:20 alle 14:20) siano presenti molti errori di ricostruzione, la media del periodo è riportata nella tabella 3.26.

Mentre dai record 14 a 20 sono presenti 3 errori molto marcati sulla media delle metriche di load (one, five, fifteen).

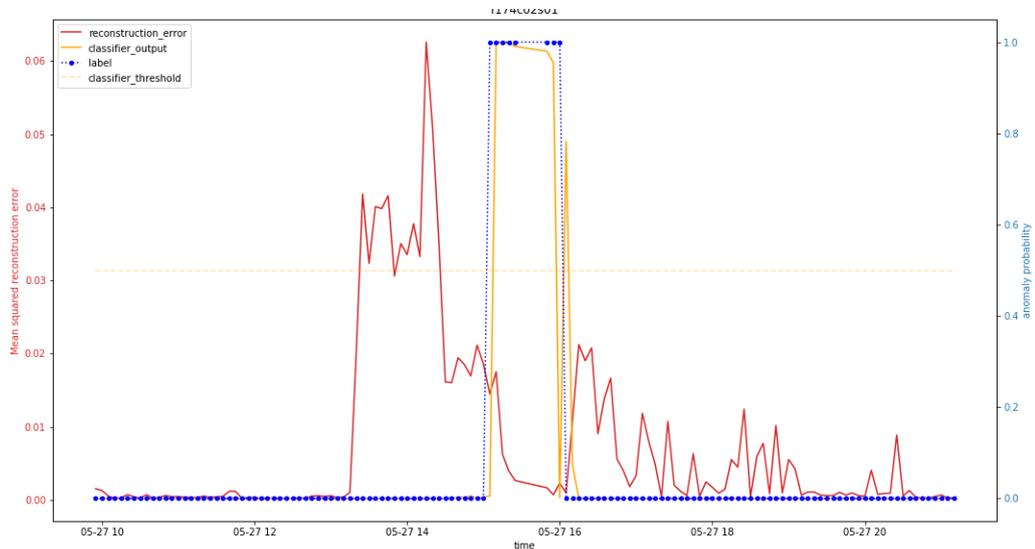


Figura 3.76: Comparazione AE-supervised AE-semisupervised r174c02s01 27 Maggio



Figura 3.77: Heatmap r174c02s01 27 Maggio

r172c02s01

Il nodo presenta l'anomalia del 27 Maggio dalle 15:20 alle 16:00 col buco di dati dalle 15:25 alle 15:50.

Nel plot (Figura 3.78) si può vedere che gli errori di ricostruzione sono concentrati attorno al periodo anomalo che sono più presenti nei periodi successivi all'anomalia.

AE-supervised ha qualche minima variazione dell'output durante l'anomalia e nei periodi successivi ma molto molto basse.

Riporto in tabella 3.27 l'errore medio durante l'anomalia.

Nel record 30 (16:20) sono presenti due errori molto elevati, entrambi sulla stessa metrica quindi sono errori relativi a media e varianza di "cpu_wio".

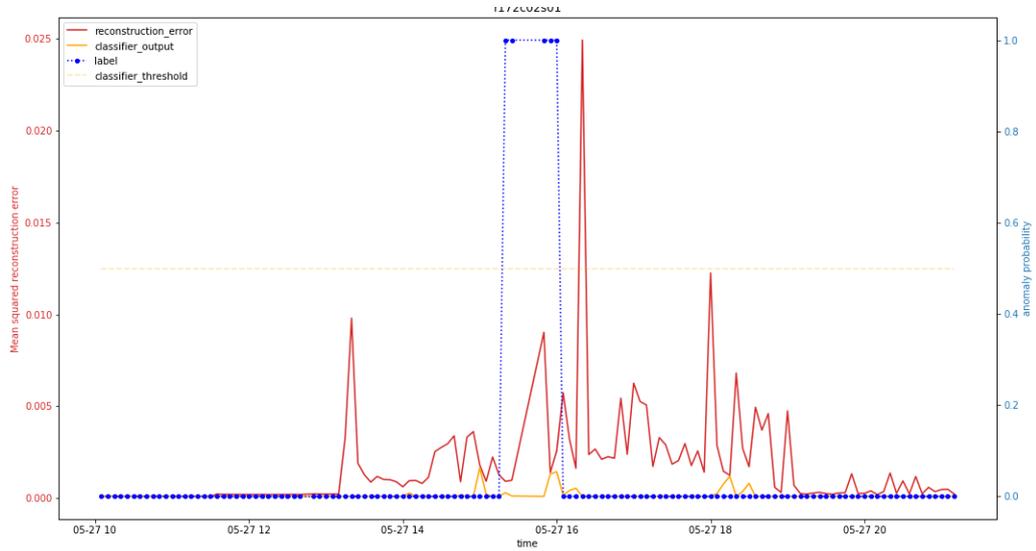


Figura 3.78: Comparazione AE-supervised AE-semisupervised r172c02s01 27 Maggio

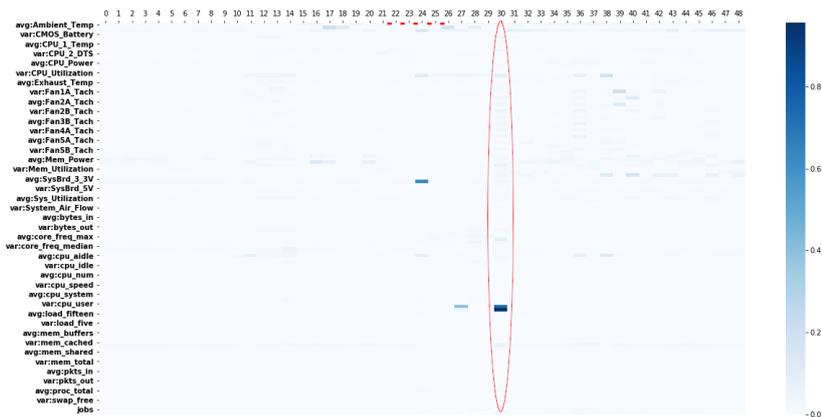


Figura 3.79: Heatmap r172c02s01 27 Maggio

var:SysBrd_3.3V	0.130896
avg:CMOS_Battery	0.052107
var:Ambient_Temp	0.036653
avg:DC_Energy	0.034873
avg:cpu_idle	0.017414
avg:Ambient_Temp	0.006768
avg:mem_free	0.005259
avg:core_freq_max	0.004542
avg:proc_total	0.004253
avg:CPU_2_DTS	0.003786

Tabella 3.27: Errore medio durante l'anomalia del 27 Maggio di r172c02s01

r136c03s01

Anche in questo nodo è presente l'anomalia del 27 Maggio dalle 15:20 alle 15:25, quindi solamente due record. È inoltre presente il buco di dati dalle 15:25 alle 15:50.

Il plot (Figura 3.80) e come analizzerò dopo anche nella heatmap presenta errori di ricostruzione frequenti su più features nelle parti precedenti all'anomalia. Come si può vedere successivamente all'anomalia l'errore cala con un singolo picco evidente.

AE-supervised invece cambia il proprio output dalle 15:05 superando anche la soglia, dando quindi un falso positivo che rileva in anticipo l'anomalia.

Nell'heatmap (Figura 3.81 ho evidenziato il periodo anomalo nel quale sono presenti la maggior parte degli errori, in tabella 3.28 ho riportato l'errore medio durante il periodo. Da notare inoltre che è sempre marcato l'errore su più features sul record 26 che rappresenta l'istante temporale delle 15:50.

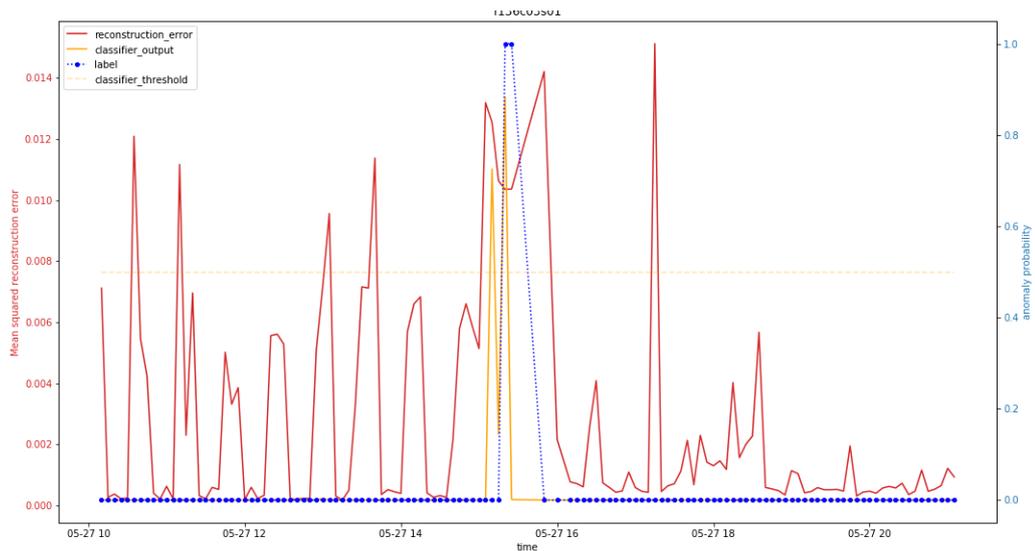


Figura 3.80: Comparazione AE-supervised AE-semisupervised r136c03s01 27 Maggio

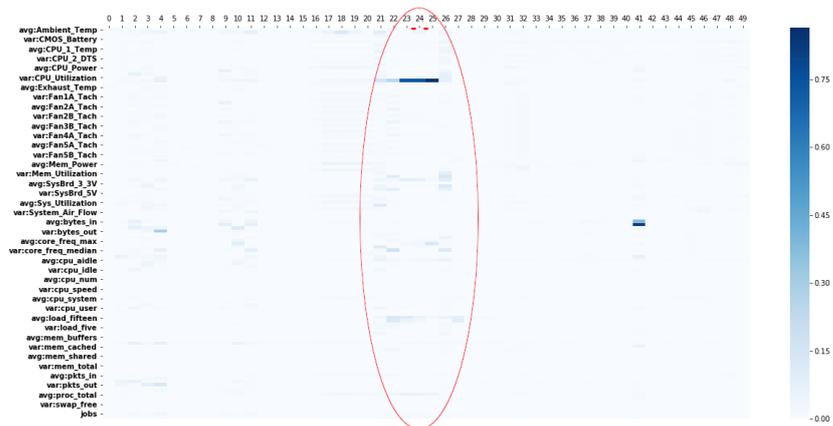


Figura 3.81: Heatmap r136c03s01 27 Maggio

avg:DC_Energy	0.804438
var:core_freq_max	0.083495
avg:proc_total	0.051764
var:SysBrd_12V	0.047656
avg:load_fifteen	0.040228
avg:cpu_idle	0.031431
var:Ambient_Temp	0.020576
var:load_fifteen	0.019866
avg:mem_cached	0.016623
avg:core_freq_max	0.012947

Tabella 3.28: Errore medio durante il periodo anomalo del 27 Maggio di r136c03s01

r130c12s03

Il nodo presenta due anomalie una il 27 Maggio e una il 25 Maggio.

Analizzo l'anomalia del 27 Maggio che parte dalle 15:20 fino alle 15:25 col buco di dati dalle 15:25 alle 15:50.

Dal plot (Figura 3.82) si nota che da circa le 14:30 l'errore di ricostruzione medio ha un trend di crescita fino a terminare in un picco alle 15:50. Da notare inoltre due spike tra le 19 e le 20 nel quale anche AE-supervised cambia il proprio output anche se di poco.

AE-supervised riesce a identificare in anticipo l'anomalia anche se inizialmente con un output minore della soglia che viene superata solamente durante il periodo dell'anomalia.

Come si può notare nell'heatmap 3.83 sono presenti molti errori nel periodo successivo all'anomalia. Durante l'anomalia l'errore più alto si ha sulla metrica della frequenza massima della cpu. Interessante come il record 26 che fa riferimento alle 15:50 ha sempre errori sparsi su più metriche. Risulta anche interessante il periodo dal record 27 al 30 (15:55 - 16:10) il quale errore è riportato in tabella 3.29.

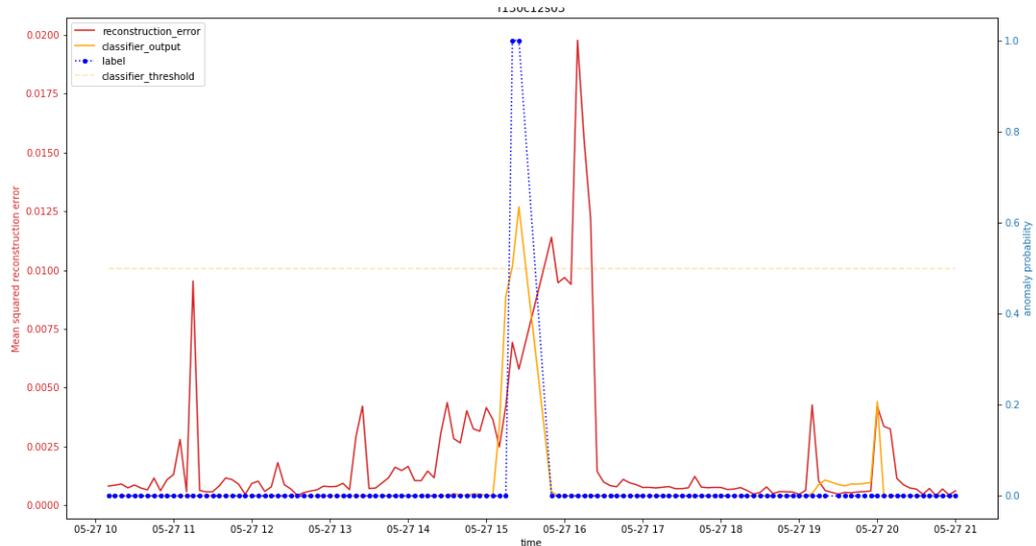


Figura 3.82: Comparazione AE-supervised AE-semisupervised r130c12s03 27 Maggio



Figura 3.83: Heatmap r130c12s03 27 Maggio

avg:CMOS_Battery	0.487704
avg:core_freq_min	0.254845
var:cpu_idle	0.204173
avg:load_fifteen	0.034638
var:cpu_user	0.033575
var:cpu_idle	0.032666
avg:cpu_idle	0.031872
avg:Sys_Utilization	0.028772
var:CPU_Utilization	0.027731
var:load_fifteen	0.026584

Tabella 3.29: Errore dalle 15:55 alle 16:10 di r130c12s03

La seconda anomalia inizia il 25 Maggio alle 20:50 fino al 26 Maggio alle 6:00.

La parte più interessante dall'anomalia a livello di heatmap è sicuramente l'inizio e riporterò solo quel grafico.

Durante l'inizio come si può vedere nel plot (Figura 3.84) si ha uno spike nell'errore di ricostruzione medio da prima delle 18:00 che successivamente mantiene un valore costante di errore con ulteriori picchi prima che l'anomalia incominci. Durante il periodo anomalo ci sono inoltre alcuni picchi di errore di ricostruzione.

AE-supervised riesce ad identificare correttamente l'anomalia (con un output sopra la soglia) tranne in un punto dove ha un drop significativo.

Nell'heatmap 3.85 si può vedere come il record 24 (20:45) presenti già errori di ricostruzione su multiple features e dall'inizio dell'anomalia inizino a comparire errori di ricostruzione, in particolare su "avg:core_freq_max" che è presente su tutta l'anomalia.

Dalle 19:15 fino alle 19:50 (dal record 6 al record 13) è molto marcato un errore su "avg:CMOS_Battery" e "jobs".

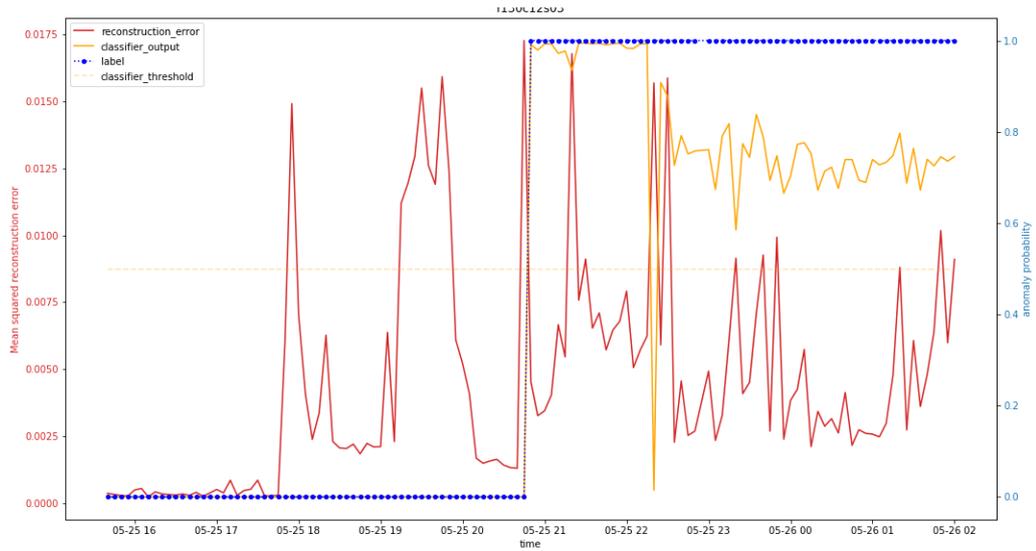


Figura 3.84: Comparazione AE-supervised AE-semisupervised r130c12s03 25 Maggio - Inizio Anomalia



Figura 3.85: Heatmap r130c12s03 25 Maggio - Inizio Anomalia

r183c05s03

Il nodo in questione presenta l'anomalia del 27 Maggio dalle 15:05 alle 16:00 con il buco di dati dalle 15:25 alle 15:50.

Nel plot si possono notare due picchi in particolare dalle 10 alle 12 e successivamente un livello di errore basso. L'errore subisce un incremento non appena inizia l'anomalia, quindi senza mostrare un aspetto anticipativo. L'errore ha un calo durante l'anomalia e risale dopo il buco di dati, subendo un incremento verso le 18

Anche AE-supervised rileva l'anomalia al primo record con un output superiore alla soglia, successivamente ha un calo nel suo output che rimane però diverso da 0 fino alle 18:00, quindi si può dire che in quel periodo anche AE-supervised rileva qualche comportamento fuori regime.

Analizzando l'heatmap (Figura 3.87 si può notare come durante l'anomalia inizino a presentarsi errori di ricostruzione su 3 features in particolare sulle medie di: "DC_Energy", "Sys_Brd12V" e "proc_total". È anche evidente che gli errori inizino a essere più frequenti dopo l'inizio dell'anomalia.

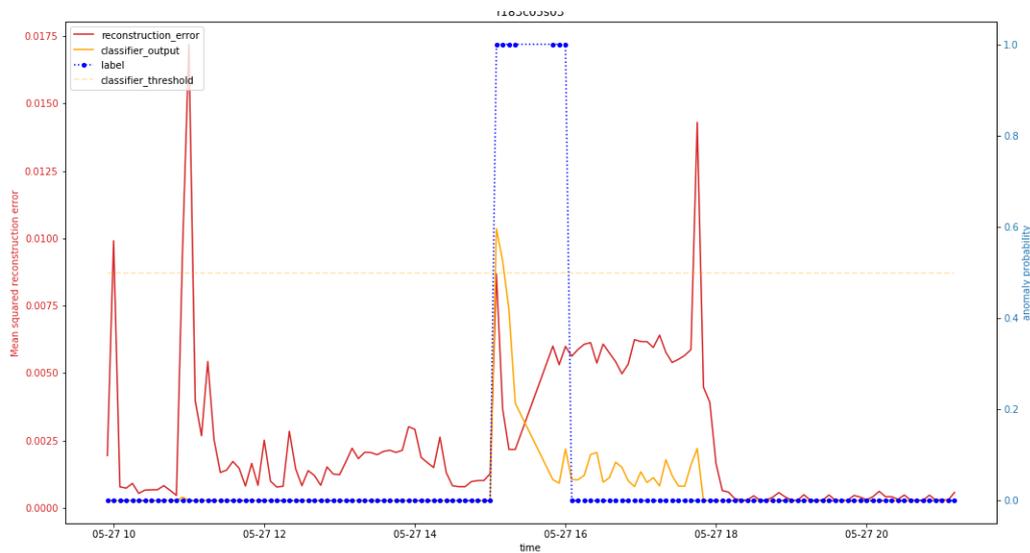


Figura 3.86: Comparazione AE-supervised AE-semisupervised r183c05s03

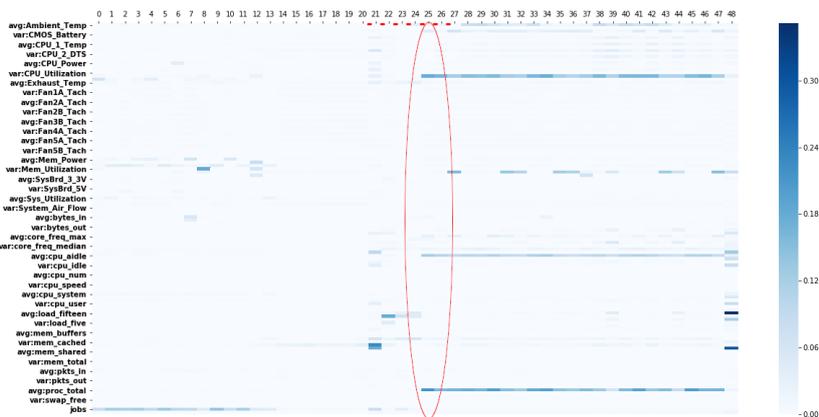


Figura 3.87: Heatmap r183c05s03 27 Maggio

r162c05s03

Il nodo presenta molteplici anomalie, analizzando le heatmap risulta interessante da analizzare l'anomalia del 27 Maggio.

Dal plot (Figura 3.88) si nota un singolo picco nell'errore alle 15:50, anche AE-supervised ha un picco nel suo output proprio alle 15:50, mentre nei primi due record anomali entrambi hanno un output basso.

Il picco dell'errore è evidente anche nell'heatmap (Figura 3.89), inoltre a

questo si può notare che successivamente all'anomalia diventa più evidente l'errore di ricostruzione su "avg:DC_Energy".

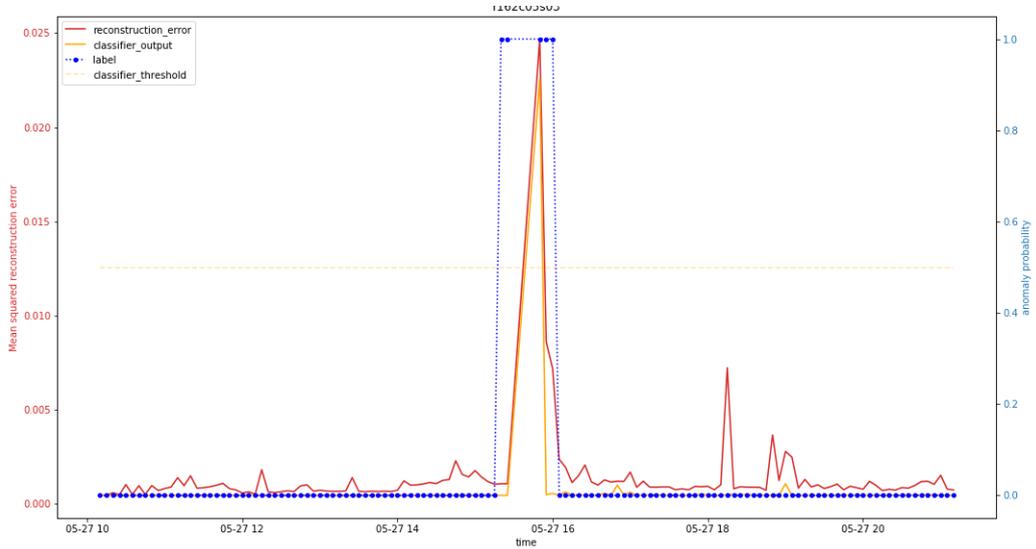


Figura 3.88: Comparazione AE-supervised AE-semisupervised r162c05s03

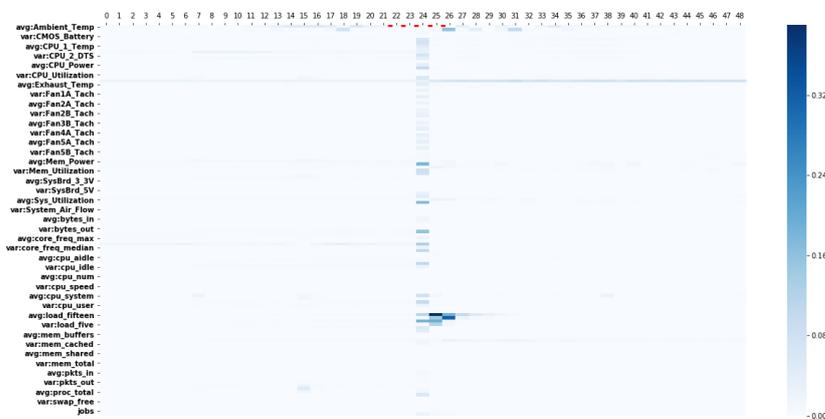


Figura 3.89: Heatmap r162c05s03 27 Maggio

r162c05s02

Questo nodo presenta la stessa anomalia di "r162c05s03" con la stessa heatmap, molto probabilmente perché avendo un codice simile sono anche

nello stesso rack e quindi c'è un'alta probabilità che si riscontrino gli stessi problemi.

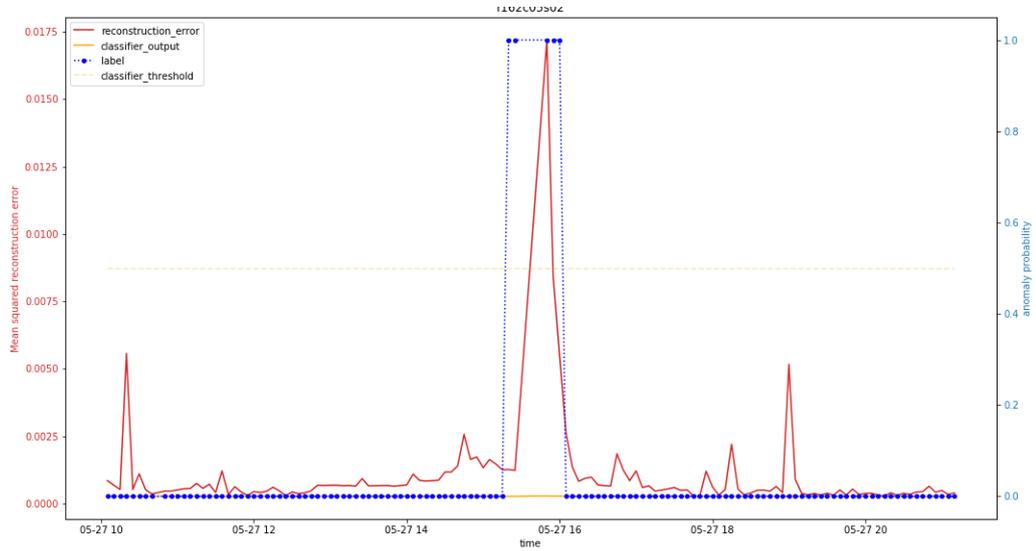


Figura 3.90: Comparazione AE-supervised AE-semisupervised r162c05s02

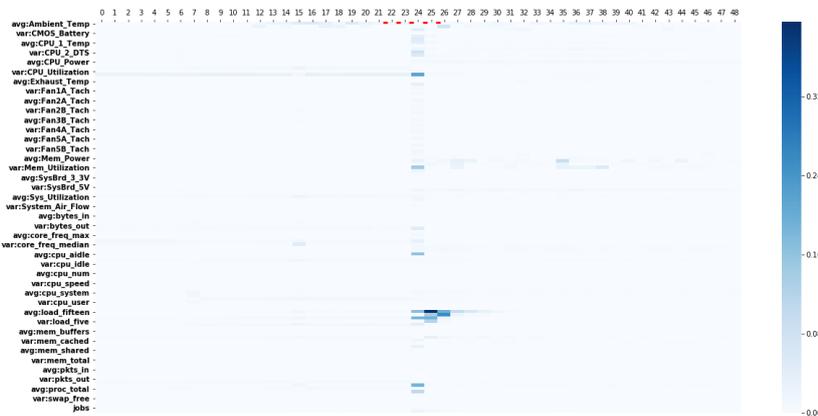


Figura 3.91: Heatmap r162c05s02

È difficile riassumere in poche parole le performance ottenute data la vastità dei dataset analizzati, è però evidente come i risultati sul periodo i Gennaio siano migliori (per ora a livello visivo) rispetto a quelli di Maggio.

Un'altra osservazione che si può fare è come i modelli analizzati siano in realtà complementari. In particolare AE-semisupervised è molto sensibile ai transitori da stato non anomalo ad anomalo mentre AE-supervised (data anche la natura della sua fase di training) è più sensibile al periodo anomalo stesso. L'ipotesi che verrà esplorata nella seguente sezione è quella di un modello ensembled.

Questa prima analisi non fornisce risultati precisi e veramente confrontabili come le metriche di performance accuracy, recall, precision, f1_score che fanno riferimento a numeri. Soprattutto non sappiamo cosa avviene nei periodi lontano da anomalie, che saranno analizzate nella prossima sezione.

3.3 Modello ensembled e detection

In questa sezione analizzerò o meglio esplorerò la possibilità di effettuare anomaly detection e soprattutto anomaly prediction tramite l'utilizzo di un modello ensembled formato da AE-supervised e AE-semisupervised.

3.3.1 Detection

Partendo da AE-supervised, dopo aver analizzato i grafici, abbiamo visto che l'output di AE-supervised è da tenere in considerazione tutte le volte che differisce da 0, dato il bias presente nel dataset e l'allenamento supervised.

Ricordando che l'output del modello supervised è la probabilità che l'istanza presa in considerazione sia anomala, porre una soglia sullo 0 significherebbe che non appena sia presente una piccola probabilità che il record sia anomalo lo consideriamo come tale. Questo sarebbe un comportamento poco tollerante nei rispetto ai possibili errori che il modello può effettuare. Abbiamo quindi deciso di usare come soglia 0.2, ovvero se il record ha come probabilità di essere anomalo maggiore del 20% allora lo consideriamo come tale.

Il modello supervised, come detto in precedenza, aiuterà il modello ensembled a rilevare l'anomalia durante il periodo anomalo e in qualche occasione a rilevarla in anticipo, come analizzato nei grafici della sezione precedente.

La fase di detection di AE-semisupervised non può essere trattata con una soglia statica poiché il suo output non è una probabilità, non ha quindi un range di valori fisso, inoltre l'errore presenta caratteristiche oscillatorie molto evidenti.

L'errore di ricostruzione si presenta infatti come un segnale con molto rumore, con oscillazioni repentine, abbiamo quindi deciso di effettuare uno smoothing di questo segnale. Lo smoothing è stato effettuato tramite la convoluzione con un filtro di media 1-D di grandezza 24.

Abbiamo quindi deciso di utilizzare una soglia variabile, in particolare questa soglia è definita dall'exponential moving average (EMA) dell'errore medio, in questo modo è possibile avere una soglia variabile che ha come riferimento l'andamento dell'errore nei periodi precedenti dando più peso agli ultimi valori osservati.

Nell'EMA è presente un parametro alfa che determina quanto peso dare all'ultimo record visto e relativamente quanto peso viene dato ai record visti nei periodi precedenti. Alfa può variare da 0 a 1 solitamente esclusi poiché ci si ritroverebbe in casi particolari che non hanno un utilizzo in questo esperimento.

Al fine di analizzare il comportamento del sistema al variare di alfa gli esperimenti sono stati effettuati con alfa pari a 0.3, 0.5, 0.7 e 0.9.

$$EMA_{t+1} = EMA_t + \alpha(Error_t + 1 - EMA_t) \quad (3.1)$$

I due modelli (AE-supervised e AE-semisupervised) sono in "or" tra loro, basta quindi che uno solo dei due modelli rilevi l'anomalia che l'output del modello ensembled sarà appunto anomalia.

Per esplorare alcune alternative e effettuare uno studio il più esauriente possibile sono stati usati come errore: errore quadrato medio, errore assoluto medio (MSE e MAE).

Nelle sezioni seguenti riporterò due versioni e tre varianti della seconda versione dei modelli che sono stati sperimentati: v1, v2-mean e v2-window mean e v2-WM&sum

3.3.2 Analisi modello ensembled v1

In questa sezione riporterò le tabelle contenenti i risultati ottenuti applicando la detection descritta in precedenza. Le colonne presenti nelle tabelle rappresentano le seguenti metriche:

- **tp**: True Positives, i record anomali che vengono identificati correttamente.
- **tn**: True Negatives, i record non anomali identificati correttamente.
- **fp_pre**: False Positives precedenti all'anomalia, sono i falsi positivi presenti nei 24 record precedenti all'inizio dell'anomalia.
- **fp_post**: False Positives successivi all'anomalia, sono i falsi positivi presenti nei 24 record successivi alla fine dell'anomalia.
- **fp_random**: Falsi positivi non nell'intorno di un'anomalia, quindi probabilmente causati dal comportamento del modello.
- **fn**: Falsi negativi, ovvero record anomali non correttamente identificati dal modello.
- **total_records**: Record totali presenti all'interno del dataset.

Sono state scelte queste metriche come indici di performance del modello perché le metriche classiche, quali accuracy, recall, precision, f1_score, non avrebbero calzato bene col nostro obiettivo di capire quanto prima riusciamo

a prevedere un'anomalia. Anzi queste metriche potevano fornire informazioni non in linea con la reale performance del modello.

Saranno presenti anche plot delle anomalie con le soglie descritte precedentemente e plot nei quali per semplicità presento l'output finale del modello ensemble, in modo da poter osservare le performance del modello tramite grafici.

Oltre alle tabelle con le metriche descritte in precedenza, inserirò il valore medio in termini di tempo di anticipo, in modo da poter quantificare quanto prima realmente in termini di tempo il modello riesce a identificare l'anomalia.

Per calcolare questa metrica abbiamo deciso di non considerare un singolo valore con output "anomalia" prima che l'anomalia accada, ma che ce ne siano almeno 3 consecutivi in modo tale da poter affermare con una certa confidenza che abbiamo rilevato un'anomalia. I valori sono riportati alla fine di ogni tabella.

Analisi Gennaio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3469.0	46.0	24.0	3376.0	0.0	6924.0
r140c12s04	81.0	3574.0	58.0	12.0	3272.0	0.0	6997.0
r174c02s02	40.0	3573.0	64.0	23.0	3270.0	5.0	6975.0
r173c10s01	15.0	3507.0	26.0	19.0	3391.0	0.0	6958.0
r130c12s04	6.0	3562.0	22.0	3.0	3402.0	0.0	6995.0
r162c05s02	70.0	3465.0	55.0	34.0	3307.0	0.0	6931.0
r129c12s02	11.0	3578.0	42.0	8.0	3310.0	1.0	6950.0
r183c05s01	6.0	781.0	0.0	15.0	737.0	0.0	1518.0
r129c12s03	14.0	3515.0	61.0	15.0	3359.0	0.0	6964.0
r141c09s01	9.0	3504.0	66.0	19.0	3390.0	3.0	6991.0
r135c09s03	22.0	3540.0	49.0	20.0	3288.0	1.0	6899.0
r162c05s03	71.0	3469.0	48.0	21.0	3315.0	2.0	6926.0
r136c04s04	20.0	3571.0	34.0	17.0	3315.0	1.0	6936.0
r183c05s02	6.0	1373.0	21.0	15.0	1217.0	0.0	2632.0
r174c02s01	3.0	1345.0	19.0	9.0	1239.0	0.0	2615.0
r172c02s01	3.0	740.0	0.0	12.0	789.0	0.0	1529.0
r136c03s01	8.0	3528.0	46.0	13.0	3327.0	0.0	6922.0
r130c12s03	12.0	3505.0	67.0	32.0	3359.0	3.0	6978.0
r183c05s03	6.0	1281.0	19.0	16.0	1303.0	0.0	2625.0
mean	21.7	2888.4	39.1	17.2	2735.1	0.8	5698.2

Continua nella prossima pagina

tp	tn	fp_pre	fp_post	fp_random	fn	records
----	----	--------	---------	-----------	----	---------

Tabella 3.30: Risultati Gennaio v1 con errore quadrato medio e alfa 0.3

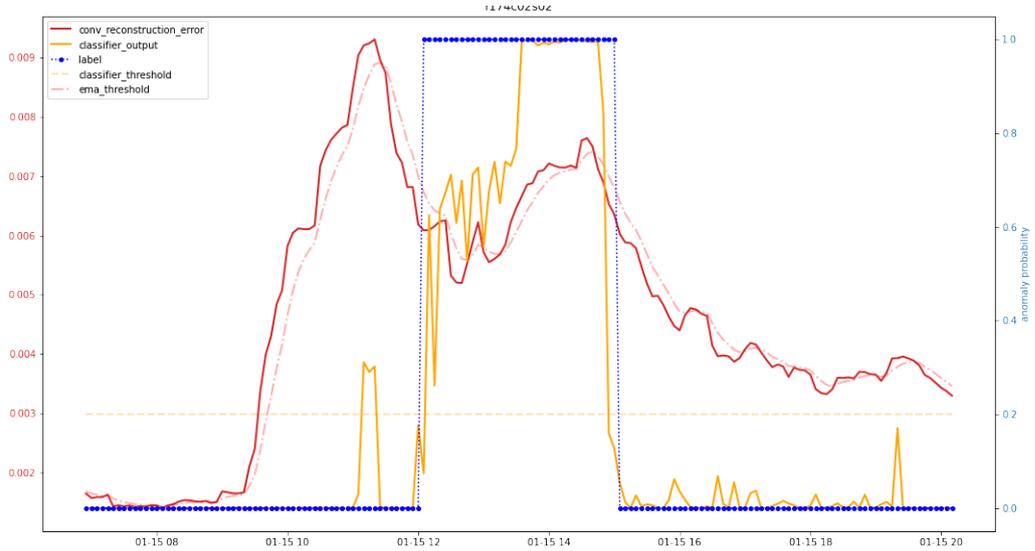


Figura 3.92: r174c02s02 anomalia 15 Gennaio, alfa 0.3

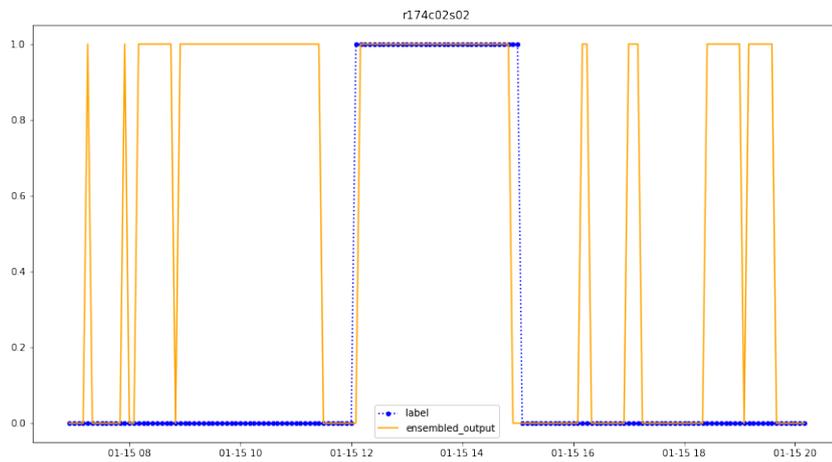


Figura 3.93: Output ensemble model anomalia 15 Gennaio, alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3463.0	43.0	23.0	3386.0	0.0	6924.0
r140c12s04	81.0	3533.0	59.0	12.0	3312.0	0.0	6997.0
r174c02s02	40.0	3541.0	62.0	27.0	3300.0	5.0	6975.0
r173c10s01	15.0	3456.0	28.0	24.0	3435.0	0.0	6958.0
r130c12s04	6.0	3516.0	23.0	4.0	3446.0	0.0	6995.0
r162c05s02	70.0	3407.0	53.0	32.0	3369.0	0.0	6931.0
r129c12s02	11.0	3547.0	44.0	9.0	3338.0	1.0	6950.0
r183c05s01	6.0	785.0	0.0	16.0	733.0	0.0	1518.0
r129c12s03	13.0	3488.0	60.0	17.0	3385.0	1.0	6964.0
r141c09s01	9.0	3519.0	63.0	24.0	3373.0	3.0	6991.0
r135c09s03	22.0	3468.0	52.0	20.0	3357.0	1.0	6899.0
r162c05s03	70.0	3474.0	46.0	21.0	3312.0	3.0	6926.0
r136c04s04	20.0	3507.0	34.0	17.0	3379.0	1.0	6936.0
r183c05s02	6.0	1363.0	20.0	15.0	1228.0	0.0	2632.0
r174c02s01	3.0	1312.0	18.0	9.0	1273.0	0.0	2615.0
r172c02s01	3.0	737.0	0.0	11.0	792.0	0.0	1529.0
r136c03s01	8.0	3487.0	45.0	14.0	3368.0	0.0	6922.0
r130c12s03	12.0	3440.0	69.0	35.0	3419.0	3.0	6978.0
r183c05s03	6.0	1291.0	21.0	18.0	1289.0	0.0	2625.0
mean	21.6	2859.7	38.9	18.3	2762.8	0.9	5698.2

Tabella 3.31: Risultati Gennaio v1 con errore quadrato medio e alfa 0.5

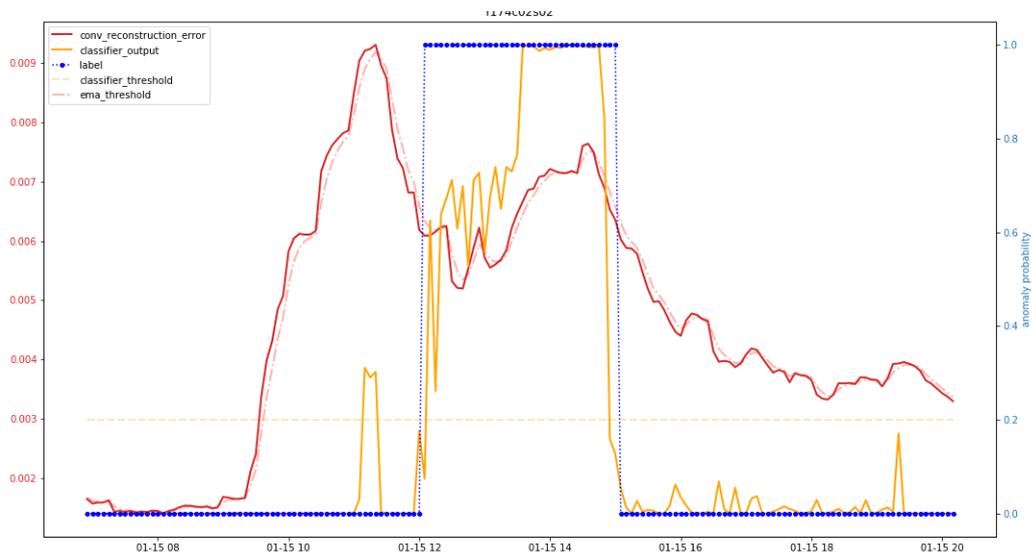


Figura 3.94: r174c02s02 anomalia 15 Gennaio, alfa 0.5

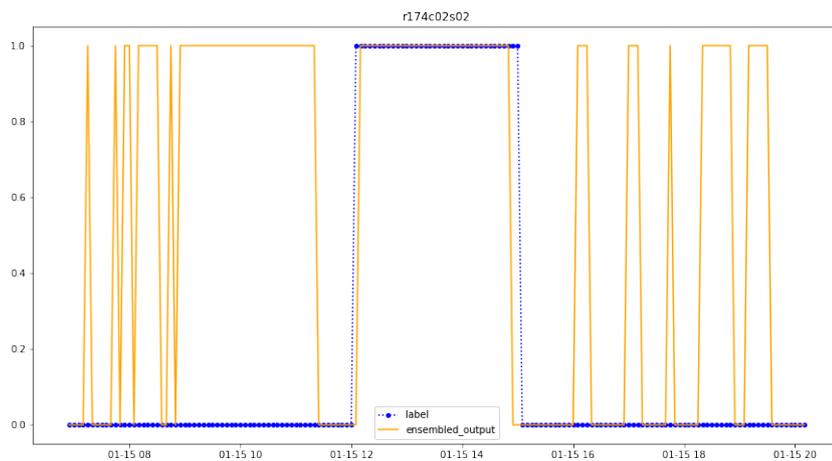


Figura 3.95: Output ensemble model anomalia 15 Gennaio, alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3460.0	43.0	22.0	3390.0	0.0	6924.0
r140c12s04	81.0	3537.0	58.0	13.0	3308.0	0.0	6997.0
r174c02s02	40.0	3504.0	60.0	25.0	3341.0	5.0	6975.0
r173c10s01	15.0	3429.0	32.0	25.0	3457.0	0.0	6958.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r130c12s04	6.0	3459.0	23.0	7.0	3500.0	0.0	6995.0
r162c05s02	70.0	3392.0	52.0	30.0	3387.0	0.0	6931.0
r129c12s02	11.0	3545.0	45.0	8.0	3340.0	1.0	6950.0
r183c05s01	6.0	782.0	0.0	16.0	736.0	0.0	1518.0
r129c12s03	13.0	3486.0	59.0	18.0	3387.0	1.0	6964.0
r141c09s01	9.0	3501.0	62.0	27.0	3389.0	3.0	6991.0
r135c09s03	22.0	3449.0	52.0	22.0	3374.0	1.0	6899.0
r162c05s03	70.0	3461.0	42.0	21.0	3329.0	3.0	6926.0
r136c04s04	20.0	3469.0	35.0	17.0	3416.0	1.0	6936.0
r183c05s02	6.0	1359.0	18.0	16.0	1233.0	0.0	2632.0
r174c02s01	3.0	1290.0	17.0	8.0	1297.0	0.0	2615.0
r172c02s01	3.0	749.0	0.0	10.0	780.0	0.0	1529.0
r136c03s01	8.0	3476.0	44.0	15.0	3379.0	0.0	6922.0
r130c12s03	12.0	3474.0	69.0	37.0	3383.0	3.0	6978.0
r183c05s03	6.0	1285.0	21.0	20.0	1293.0	0.0	2625.0
mean	21.6	2847.7	38.5	18.8	2774.7	0.9	5698.2

Tabella 3.32: Risultati Gennaio v1 con errore quadrato medio e alfa 0.7

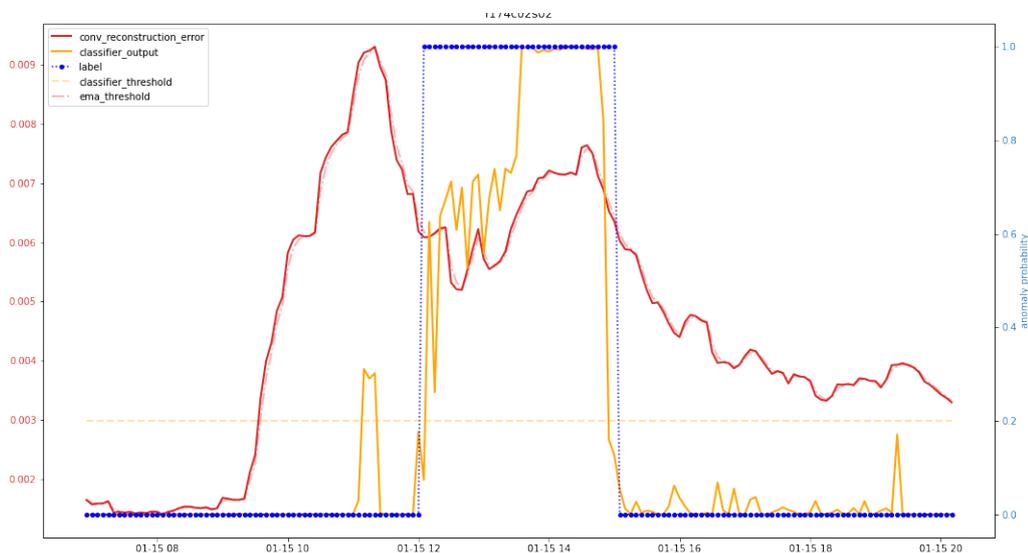


Figura 3.96: r174c02s02 anomalia 15 Gennaio, alfa 0.7

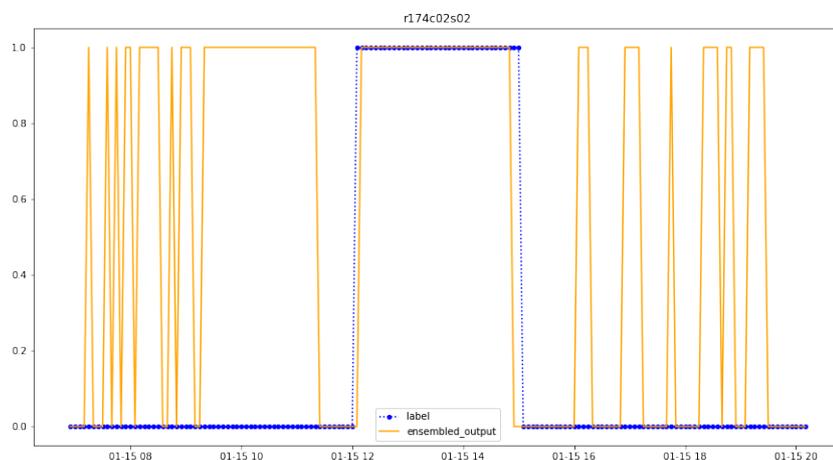


Figura 3.97: Output ensembled model anomalia 15 Gennaio, alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3429.0	44.0	25.0	3417.0	0.0	6924.0
r140c12s04	81.0	3482.0	58.0	12.0	3364.0	0.0	6997.0
r174c02s02	40.0	3519.0	54.0	23.0	3334.0	5.0	6975.0
r173c10s01	15.0	3441.0	35.0	23.0	3444.0	0.0	6958.0
r130c12s04	6.0	3469.0	24.0	8.0	3488.0	0.0	6995.0
r162c05s02	70.0	3371.0	54.0	28.0	3408.0	0.0	6931.0
r129c12s02	11.0	3554.0	45.0	10.0	3329.0	1.0	6950.0
r183c05s01	6.0	787.0	0.0	19.0	731.0	0.0	1518.0
r129c12s03	13.0	3471.0	60.0	21.0	3398.0	1.0	6964.0
r141c09s01	9.0	3503.0	62.0	27.0	3387.0	3.0	6991.0
r135c09s03	22.0	3433.0	53.0	22.0	3389.0	1.0	6899.0
r162c05s03	70.0	3457.0	45.0	23.0	3328.0	3.0	6926.0
r136c04s04	20.0	3445.0	34.0	18.0	3440.0	1.0	6936.0
r183c05s02	6.0	1327.0	18.0	17.0	1264.0	0.0	2632.0
r174c02s01	3.0	1284.0	17.0	8.0	1303.0	0.0	2615.0
r172c02s01	3.0	758.0	0.0	10.0	771.0	0.0	1529.0
r136c03s01	8.0	3450.0	43.0	15.0	3406.0	0.0	6922.0
r130c12s03	12.0	3451.0	70.0	36.0	3406.0	3.0	6978.0
r183c05s03	6.0	1294.0	21.0	21.0	1283.0	0.0	2625.0
mean	21.6	2838.2	38.8	19.3	2783.7	0.9	5698.2

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
--	----	----	--------	---------	-----------	----	---------

Tabella 3.33: Risultati Gennaio v1 con errore quadrato medio e alfa 0.9

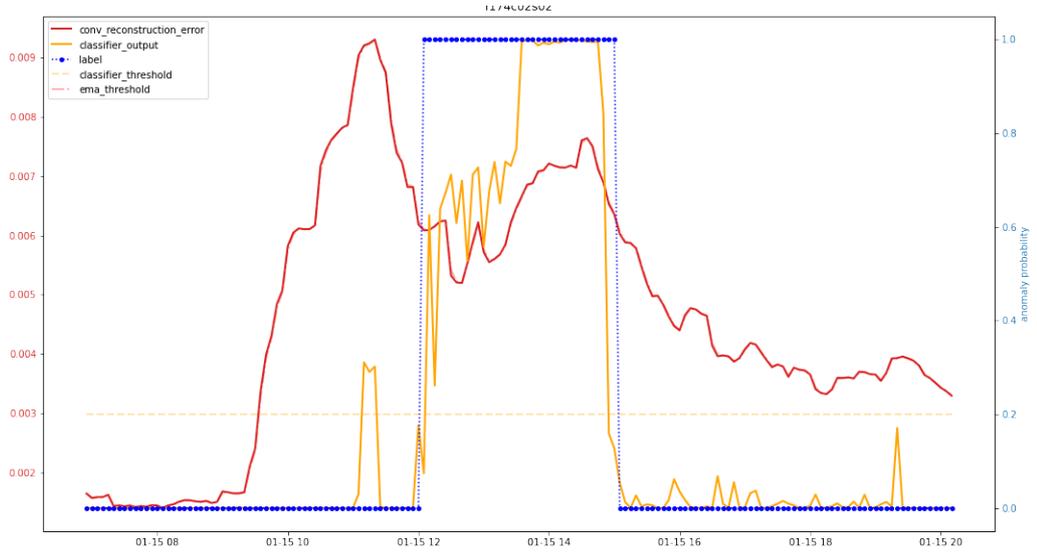


Figura 3.98: r174c02s02 anomalia 15 Gennaio, alfa 0.9

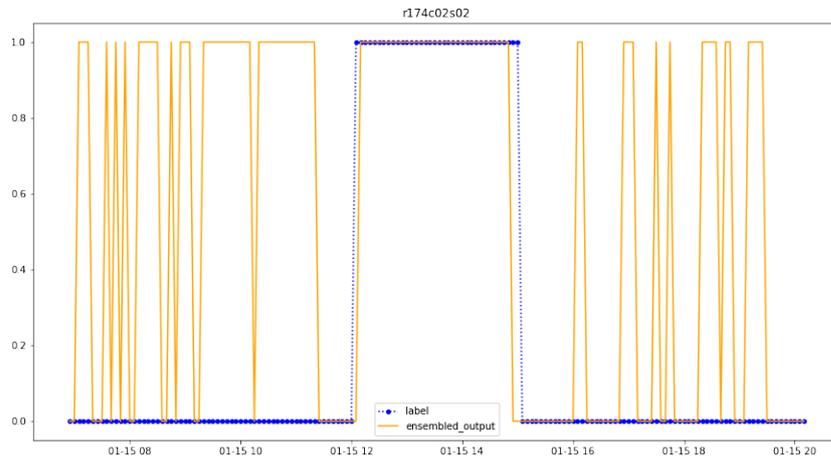


Figura 3.99: Output ensemble model anomalia 15 Gennaio, errore quadrato, alfa 0.9

Alfa	Anticipo medio
0.3	2 h 53 min
0.5	2 h 54 min
0.7	2 h 55 min
0.9	2 h 52 min

Tabella 3.34: Tabella anticipo medio Gennaio v1 - errore quadrato

Analisi Gennaio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3463.0	44.0	22.0	3386.0	0.0	6924.0
r140c12s04	81.0	3544.0	68.0	22.0	3282.0	0.0	6997.0
r174c02s02	40.0	3500.0	64.0	15.0	3351.0	5.0	6975.0
r173c10s01	15.0	3526.0	26.0	20.0	3371.0	0.0	6958.0
r130c12s04	6.0	3534.0	22.0	3.0	3430.0	0.0	6995.0
r162c05s02	64.0	3524.0	40.0	14.0	3283.0	6.0	6931.0
r129c12s02	11.0	3618.0	44.0	9.0	3267.0	1.0	6950.0
r183c05s01	6.0	750.0	0.0	17.0	768.0	0.0	1518.0
r129c12s03	13.0	3511.0	63.0	13.0	3363.0	1.0	6964.0
r141c09s01	9.0	3544.0	67.0	22.0	3346.0	3.0	6991.0
r135c09s03	22.0	3590.0	49.0	20.0	3238.0	1.0	6899.0
r162c05s03	64.0	3491.0	39.0	12.0	3311.0	9.0	6926.0
r136c04s04	18.0	3549.0	34.0	14.0	3338.0	3.0	6936.0
r183c05s02	6.0	1256.0	20.0	15.0	1335.0	0.0	2632.0
r174c02s01	3.0	1337.0	19.0	10.0	1246.0	0.0	2615.0
r172c02s01	3.0	738.0	0.0	12.0	791.0	0.0	1529.0
r136c03s01	8.0	3513.0	44.0	14.0	3343.0	0.0	6922.0
r130c12s03	12.0	3512.0	49.0	23.0	3379.0	3.0	6978.0
r183c05s03	6.0	1288.0	20.0	16.0	1295.0	0.0	2625.0
mean	20.8	2883.6	37.5	15.4	2743.3	1.7	5698.2

Tabella 3.35: Risultati Gennaio v1 con errore assoluto medio e alfa 0.3

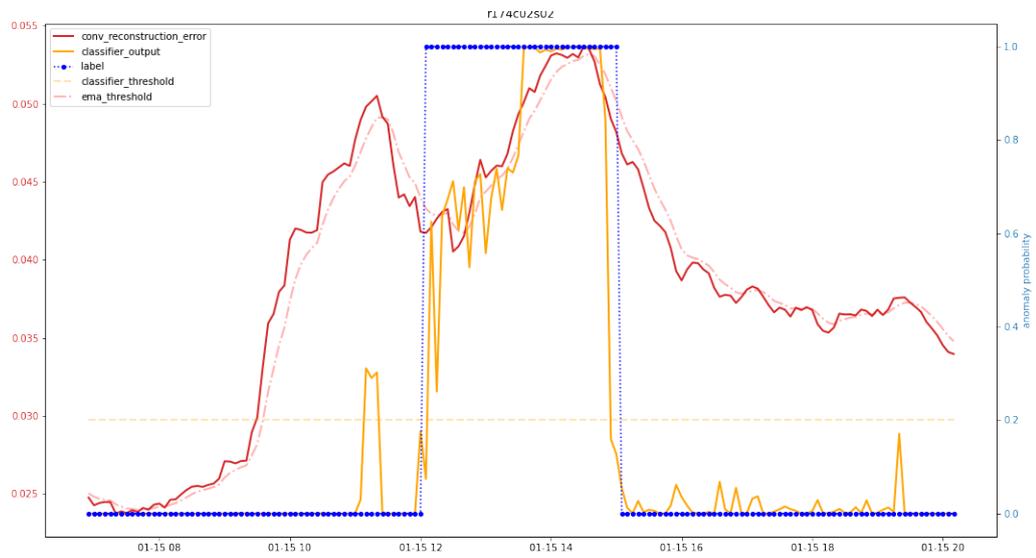


Figura 3.100: r174c02s02 anomalia 15 Gennaio, alfa 0.3

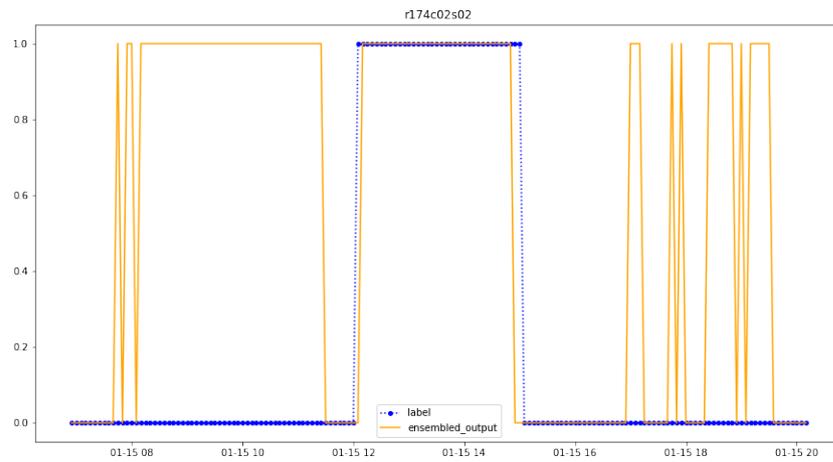


Figura 3.101: Output ensemble model anomalia 15 Gennaio, alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3463.0	45.0	23.0	3384.0	0.0	6924.0
r140c12s04	81.0	3502.0	68.0	21.0	3325.0	0.0	6997.0
r174c02s02	40.0	3451.0	59.0	17.0	3403.0	5.0	6975.0
r173c10s01	15.0	3499.0	29.0	25.0	3390.0	0.0	6958.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r130c12s04	6.0	3457.0	23.0	6.0	3503.0	0.0	6995.0
r162c05s02	64.0	3477.0	38.0	15.0	3331.0	6.0	6931.0
r129c12s02	11.0	3586.0	46.0	9.0	3297.0	1.0	6950.0
r183c05s01	6.0	747.0	0.0	18.0	771.0	0.0	1518.0
r129c12s03	13.0	3497.0	62.0	16.0	3375.0	1.0	6964.0
r141c09s01	9.0	3520.0	64.0	24.0	3371.0	3.0	6991.0
r135c09s03	22.0	3567.0	51.0	21.0	3258.0	1.0	6899.0
r162c05s03	64.0	3447.0	35.0	18.0	3353.0	9.0	6926.0
r136c04s04	19.0	3547.0	35.0	15.0	3339.0	2.0	6936.0
r183c05s02	6.0	1293.0	18.0	15.0	1300.0	0.0	2632.0
r174c02s01	3.0	1303.0	19.0	9.0	1281.0	0.0	2615.0
r172c02s01	3.0	720.0	0.0	11.0	809.0	0.0	1529.0
r136c03s01	8.0	3529.0	42.0	14.0	3329.0	0.0	6922.0
r130c12s03	13.0	3509.0	51.0	27.0	3376.0	2.0	6978.0
r183c05s03	6.0	1297.0	20.0	16.0	1286.0	0.0	2625.0
mean	20.9	2863.7	37.1	16.8	2762.2	1.6	5698.2

Tabella 3.36: Risultati Gennaio v1 con errore assoluto medio e alfa 0.5

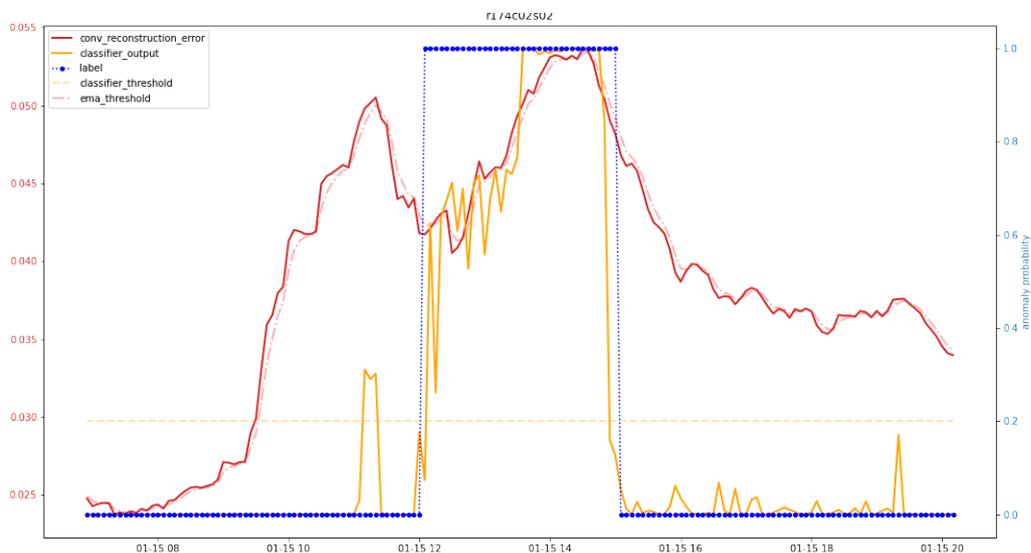


Figura 3.102: r174c02s02 anomalia 15 Gennaio, alfa 0.5

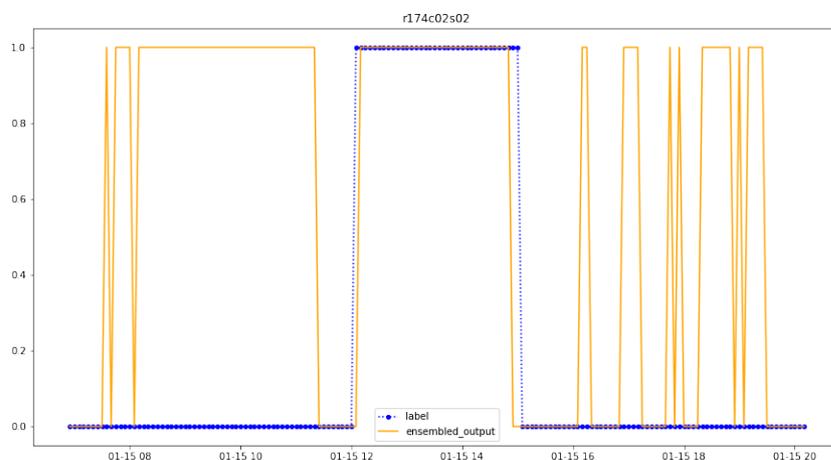


Figura 3.103: Output ensembled model anomalia 15 Gennaio, alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3444.0	44.0	23.0	3404.0	0.0	6924.0
r140c12s04	81.0	3467.0	69.0	23.0	3357.0	0.0	6997.0
r174c02s02	40.0	3429.0	57.0	18.0	3426.0	5.0	6975.0
r173c10s01	15.0	3469.0	33.0	24.0	3417.0	0.0	6958.0
r130c12s04	6.0	3426.0	23.0	7.0	3533.0	0.0	6995.0
r162c05s02	64.0	3423.0	36.0	16.0	3386.0	6.0	6931.0
r129c12s02	11.0	3579.0	46.0	8.0	3305.0	1.0	6950.0
r183c05s01	6.0	749.0	0.0	19.0	769.0	0.0	1518.0
r129c12s03	13.0	3488.0	62.0	19.0	3381.0	1.0	6964.0
r141c09s01	9.0	3506.0	63.0	26.0	3384.0	3.0	6991.0
r135c09s03	22.0	3514.0	52.0	21.0	3310.0	1.0	6899.0
r162c05s03	64.0	3449.0	35.0	19.0	3350.0	9.0	6926.0
r136c04s04	19.0	3493.0	35.0	15.0	3393.0	2.0	6936.0
r183c05s02	6.0	1281.0	19.0	15.0	1311.0	0.0	2632.0
r174c02s01	3.0	1314.0	17.0	8.0	1273.0	0.0	2615.0
r172c02s01	3.0	735.0	0.0	10.0	794.0	0.0	1529.0
r136c03s01	8.0	3526.0	40.0	15.0	3333.0	0.0	6922.0
r130c12s03	13.0	3481.0	50.0	26.0	3406.0	2.0	6978.0
r183c05s03	6.0	1279.0	20.0	16.0	1304.0	0.0	2625.0
mean	20.9	2844.8	36.9	17.3	2780.8	1.6	5698.2

Continua nella prossima pagina

tp	tn	fp_pre	fp_post	fp_random	fn	records
----	----	--------	---------	-----------	----	---------

Tabella 3.37: Risultati Gennaio v1 con errore assoluto medio e alfa 0.7

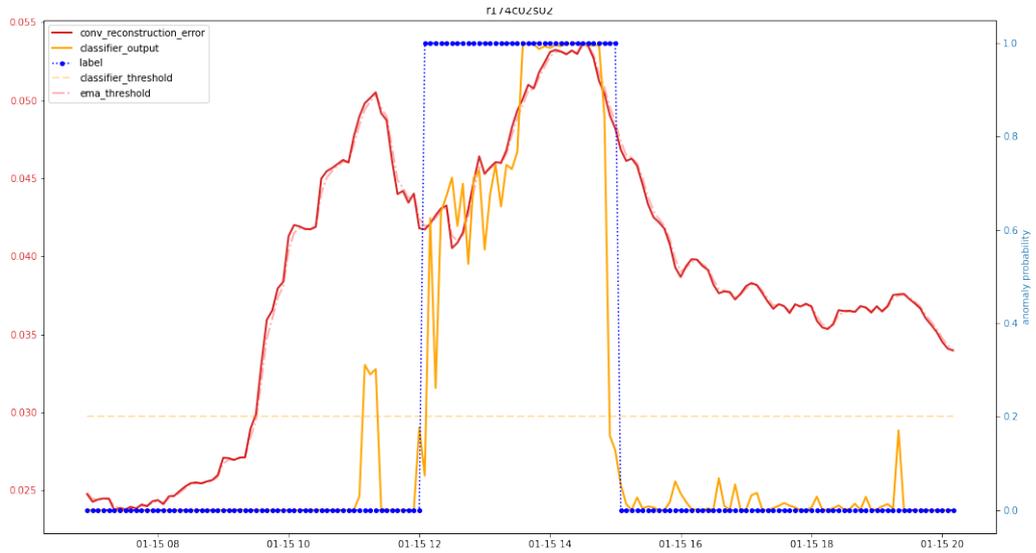


Figura 3.104: r174c02s02 anomalia 15 Gennaio, alfa 0.7

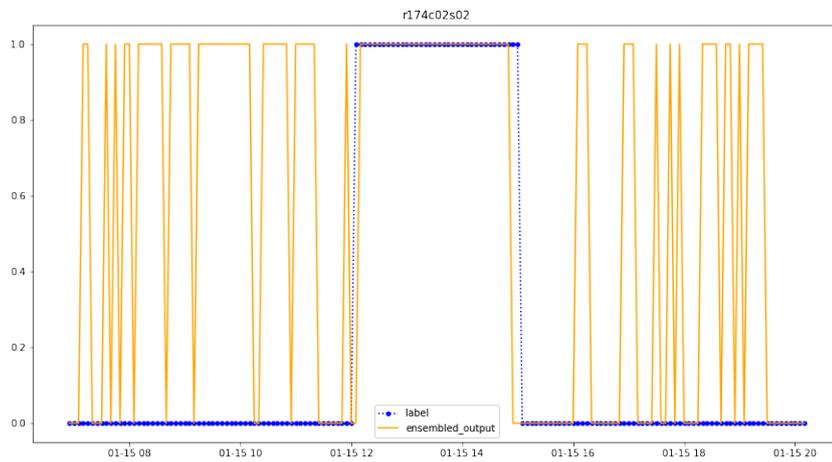


Figura 3.105: Output ensemble model anomalia 15 Gennaio, alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	3449.0	43.0	23.0	3400.0	0.0	6924.0
r140c12s04	81.0	3456.0	67.0	25.0	3368.0	0.0	6997.0
r174c02s02	40.0	3403.0	56.0	21.0	3450.0	5.0	6975.0
r173c10s01	15.0	3508.0	33.0	25.0	3377.0	0.0	6958.0
r130c12s04	6.0	3425.0	23.0	8.0	3533.0	0.0	6995.0
r162c05s02	64.0	3423.0	36.0	15.0	3387.0	6.0	6931.0
r129c12s02	11.0	3545.0	46.0	9.0	3338.0	1.0	6950.0
r183c05s01	6.0	737.0	0.0	18.0	781.0	0.0	1518.0
r129c12s03	13.0	3465.0	64.0	22.0	3399.0	1.0	6964.0
r141c09s01	10.0	3460.0	61.0	26.0	3432.0	2.0	6991.0
r135c09s03	22.0	3454.0	53.0	22.0	3368.0	1.0	6899.0
r162c05s03	65.0	3445.0	36.0	18.0	3354.0	8.0	6926.0
r136c04s04	19.0	3482.0	35.0	15.0	3404.0	2.0	6936.0
r183c05s02	6.0	1279.0	18.0	15.0	1314.0	0.0	2632.0
r174c02s01	3.0	1296.0	17.0	8.0	1291.0	0.0	2615.0
r172c02s01	3.0	741.0	0.0	10.0	788.0	0.0	1529.0
r136c03s01	8.0	3524.0	41.0	15.0	3334.0	0.0	6922.0
r130c12s03	13.0	3473.0	50.0	27.0	3413.0	2.0	6978.0
r183c05s03	6.0	1276.0	21.0	16.0	1306.0	0.0	2625.0
mean	21.1	2833.7	36.8	17.8	2791.4	1.5	5698.2

Tabella 3.38: Risultati Gennaio v1 con errore assoluto medio e alfa 0.9

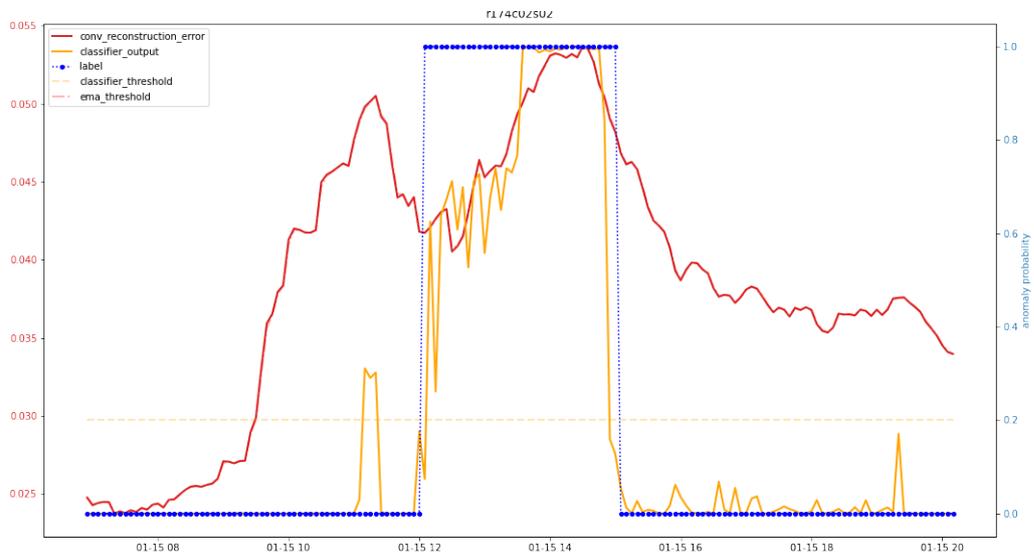


Figura 3.106: r174c02s02 anomalia 15 Gennaio, alfa 0.9

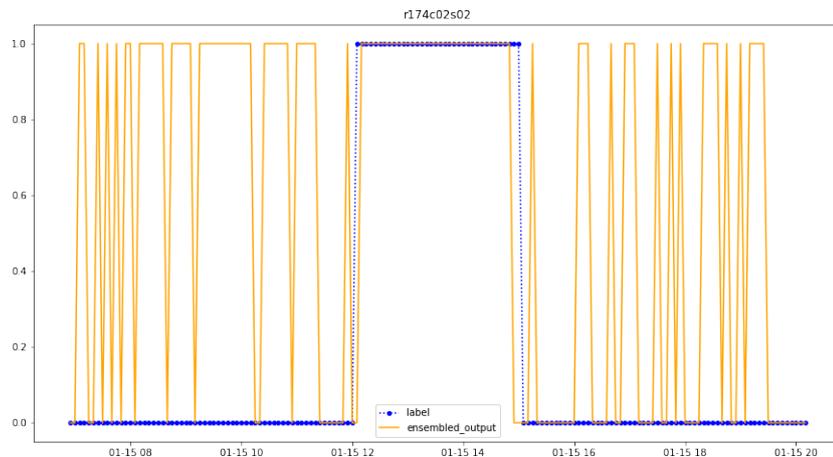


Figura 3.107: Output ensemble model anomalia 15 Gennaio, alfa 0.9

Il trend di questa metodologia inizia a essere evidente, ci sono molti falsi positivi e la presenza così alta di fp_random è un problema poiché causerebbero un falso allarme ai sysadmin di Marconi causando una grave perdita di tempo dato anche il fatto che i falsi positivi sono circa il 50% dell'intero dataset.

Inoltre si può anche notare come il variare di alfa non influenzi in modo

Alfa	Anticipo medio
0.3	2 h 48 min
0.5	2 h 51 min
0.7	2 h 49 min
0.9	2 h 50 min

Tabella 3.39: Tabella anticipo medio Gennaio v1 - errore assoluto

significativo nè random fp nè l'anticipo medio.

Riporto anche le analisi di Maggio, per evidenziare che il problema non è strettamente legato al dataset del mese di Gennaio.

Analisi Maggio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	3570.0	15.0	14.0	3514.0	3.0	7118.0
r141c09s02	2.0	3688.0	18.0	2.0	3500.0	0.0	7210.0
r140c12s04	5.0	3708.0	43.0	21.0	3432.0	0.0	7209.0
r174c02s02	5.0	3696.0	16.0	14.0	3481.0	0.0	7212.0
r173c10s01	5.0	3503.0	22.0	16.0	3609.0	0.0	7155.0
r130c12s04	111.0	3639.0	33.0	6.0	3409.0	0.0	7198.0
r162c05s02	1320.0	2716.0	142.0	112.0	2852.0	52.0	7165.0
r129c12s02	2.0	3602.0	24.0	2.0	3493.0	0.0	7123.0
r183c05s01	4.0	3606.0	13.0	13.0	3569.0	0.0	7205.0
r129c12s03	2.0	3602.0	18.0	9.0	3450.0	0.0	7081.0
r141c09s01	2.0	3681.0	24.0	2.0	3498.0	0.0	7207.0
r135c09s03	1.0	3641.0	24.0	0.0	3477.0	1.0	7144.0
r162c05s03	1363.0	2602.0	152.0	122.0	2995.0	13.0	7165.0
r136c04s04	2.0	3554.0	12.0	9.0	3578.0	0.0	7155.0
r183c05s02	864.0	3174.0	40.0	12.0	3107.0	1.0	7198.0
r174c02s01	6.0	3664.0	17.0	1.0	3523.0	2.0	7213.0
r172c02s01	5.0	3677.0	22.0	14.0	3470.0	0.0	7188.0
r136c03s01	1.0	3585.0	16.0	5.0	3515.0	1.0	7123.0
r130c12s03	112.0	3594.0	38.0	25.0	3432.0	0.0	7201.0
r183c05s03	7.0	3599.0	18.0	13.0	3572.0	0.0	7209.0
mean	191.1	3505.1	35.4	20.6	3423.8	3.6	7173.9

Tabella 3.40: Risultati Maggio v1 con errore quadrato medio e alfa 0.3

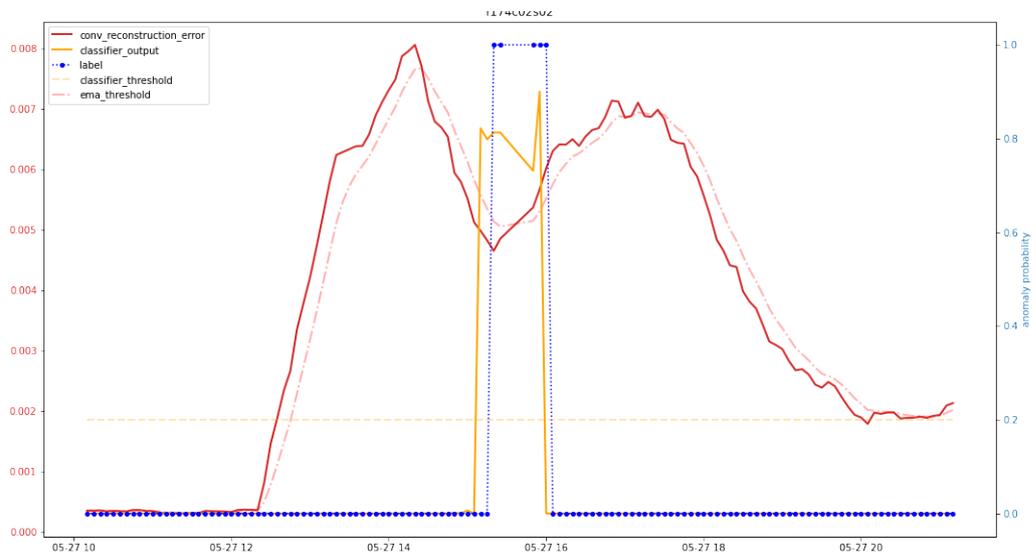


Figura 3.108: r174c02s02 anomalia 27 Maggio, alfa 0.3

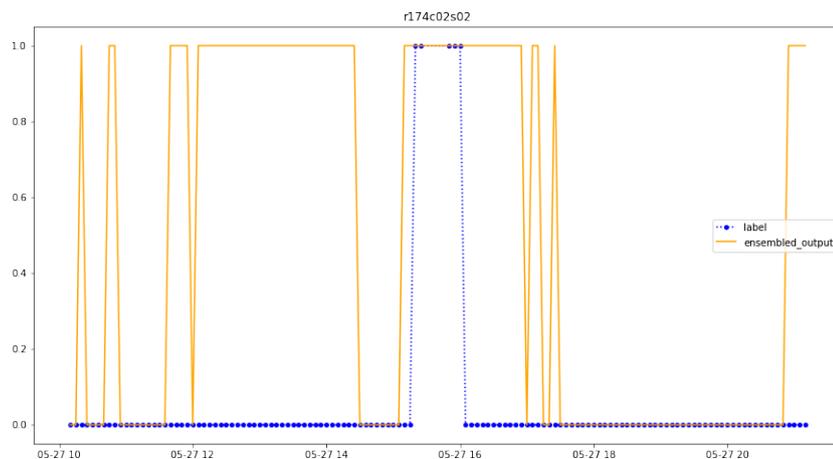


Figura 3.109: Output ensemble model anomalia 27 Maggio, alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	3.0	3576.0	13.0	15.0	3509.0	2.0	7118.0
r141c09s02	2.0	3661.0	18.0	3.0	3526.0	0.0	7210.0
r140c12s04	5.0	3683.0	42.0	22.0	3457.0	0.0	7209.0
r174c02s02	5.0	3687.0	15.0	12.0	3493.0	0.0	7212.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r173c10s01	5.0	3526.0	22.0	15.0	3587.0	0.0	7155.0
r130c12s04	111.0	3608.0	35.0	10.0	3434.0	0.0	7198.0
r162c05s02	1317.0	2759.0	145.0	109.0	2811.0	55.0	7165.0
r129c12s02	2.0	3581.0	24.0	1.0	3515.0	0.0	7123.0
r183c05s01	4.0	3543.0	15.0	12.0	3631.0	0.0	7205.0
r129c12s03	2.0	3596.0	18.0	9.0	3456.0	0.0	7081.0
r141c09s01	2.0	3637.0	24.0	4.0	3540.0	0.0	7207.0
r135c09s03	1.0	3572.0	23.0	2.0	3545.0	1.0	7144.0
r162c05s03	1360.0	2605.0	152.0	126.0	2988.0	16.0	7165.0
r136c04s04	1.0	3559.0	13.0	7.0	3574.0	1.0	7155.0
r183c05s02	864.0	3167.0	40.0	11.0	3115.0	1.0	7198.0
r174c02s01	6.0	3684.0	16.0	2.0	3503.0	2.0	7213.0
r172c02s01	5.0	3696.0	21.0	15.0	3451.0	0.0	7188.0
r136c03s01	1.0	3608.0	17.0	8.0	3488.0	1.0	7123.0
r130c12s03	112.0	3585.0	32.0	24.0	3448.0	0.0	7201.0
r183c05s03	7.0	3572.0	18.0	12.0	3600.0	0.0	7209.0
mean	190.8	3495.2	35.1	20.9	3433.6	4.0	7173.9

Tabella 3.41: Risultati Maggio v1 con errore quadrato medio e alfa 0.5

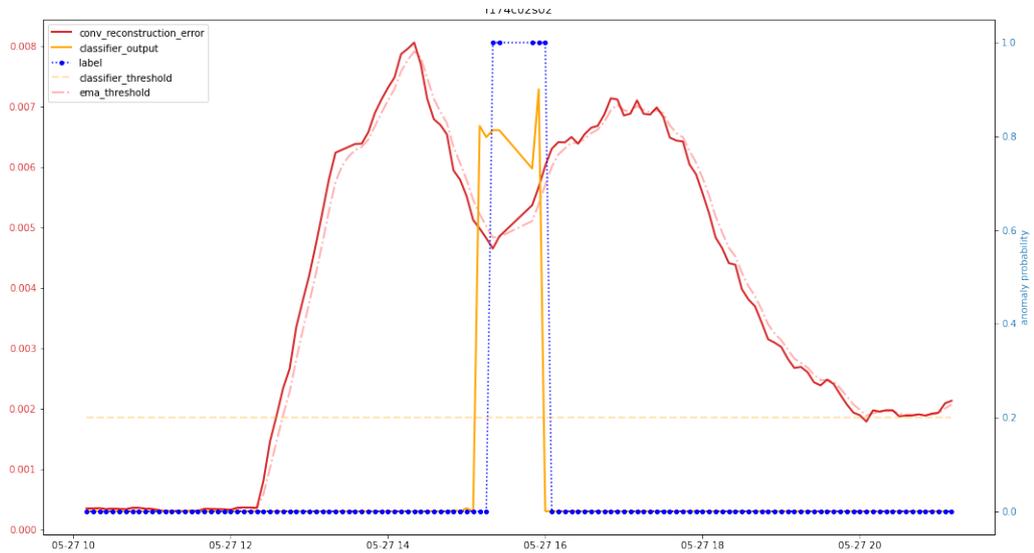


Figura 3.110: r174c02s02 anomalia 27 Maggio, alfa 0.5

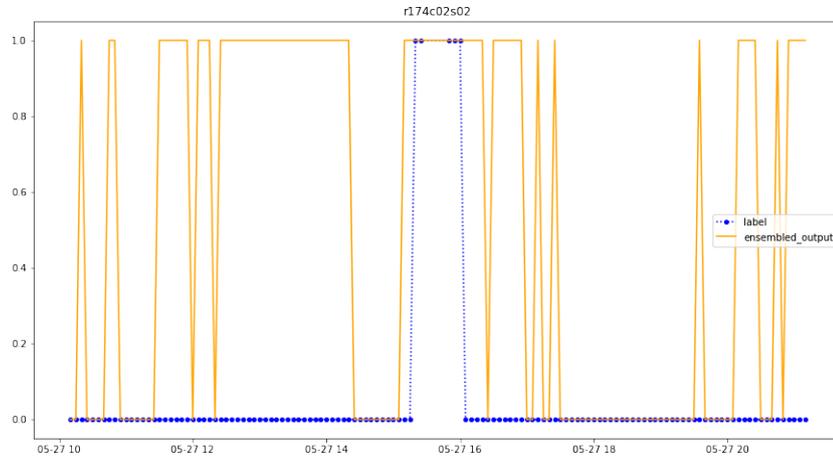


Figura 3.111: Output ensemble model anomalia 27 Maggio, alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	4.0	3592.0	13.0	15.0	3493.0	1.0	7118.0
r141c09s02	2.0	3634.0	18.0	4.0	3552.0	0.0	7210.0
r140c12s04	5.0	3703.0	41.0	24.0	3436.0	0.0	7209.0
r174c02s02	5.0	3706.0	15.0	12.0	3474.0	0.0	7212.0
r173c10s01	5.0	3535.0	22.0	16.0	3577.0	0.0	7155.0
r130c12s04	110.0	3567.0	33.0	9.0	3478.0	1.0	7198.0
r162c05s02	1319.0	2720.0	144.0	109.0	2851.0	53.0	7165.0
r129c12s02	2.0	3567.0	24.0	2.0	3528.0	0.0	7123.0
r183c05s01	4.0	3544.0	15.0	11.0	3631.0	0.0	7205.0
r129c12s03	2.0	3561.0	18.0	9.0	3491.0	0.0	7081.0
r141c09s01	1.0	3630.0	22.0	5.0	3548.0	1.0	7207.0
r135c09s03	1.0	3569.0	23.0	2.0	3548.0	1.0	7144.0
r162c05s03	1360.0	2589.0	153.0	126.0	3004.0	16.0	7165.0
r136c04s04	1.0	3552.0	12.0	5.0	3584.0	1.0	7155.0
r183c05s02	864.0	3146.0	41.0	10.0	3136.0	1.0	7198.0
r174c02s01	6.0	3674.0	16.0	5.0	3510.0	2.0	7213.0
r172c02s01	5.0	3688.0	21.0	13.0	3461.0	0.0	7188.0
r136c03s01	1.0	3612.0	14.0	10.0	3485.0	1.0	7123.0
r130c12s03	112.0	3612.0	31.0	24.0	3422.0	0.0	7201.0
r183c05s03	7.0	3551.0	16.0	10.0	3625.0	0.0	7209.0
mean	190.8	3487.6	34.6	21.1	3441.7	3.9	7173.9

Continua nella prossima pagina

tp	tn	fp_pre	fp_post	fp_random	fn	records
----	----	--------	---------	-----------	----	---------

Tabella 3.42: Risultati Maggio v1 con errore quadrato medio e alfa 0.7

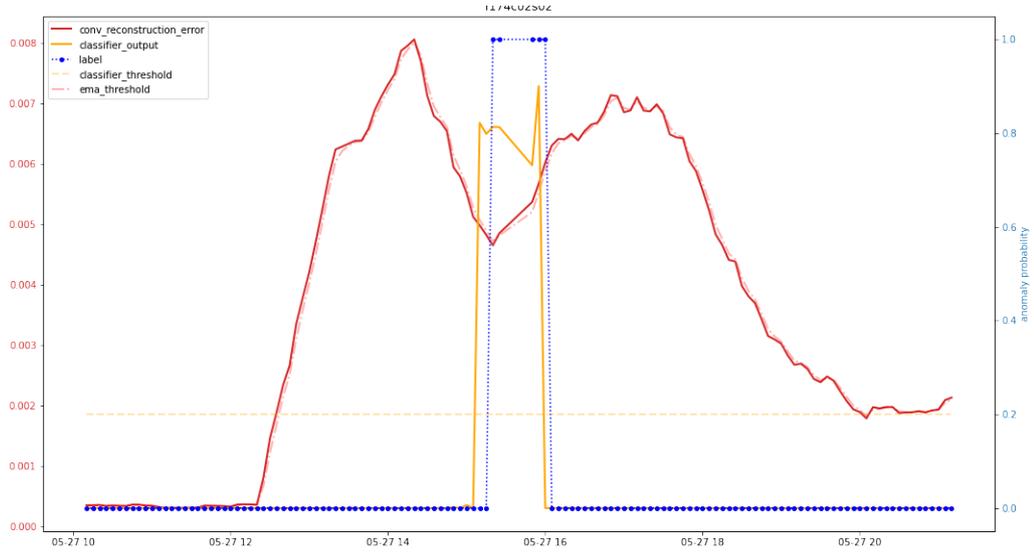


Figura 3.112: r174c02s02 anomalia 27 Maggio, alfa 0.7

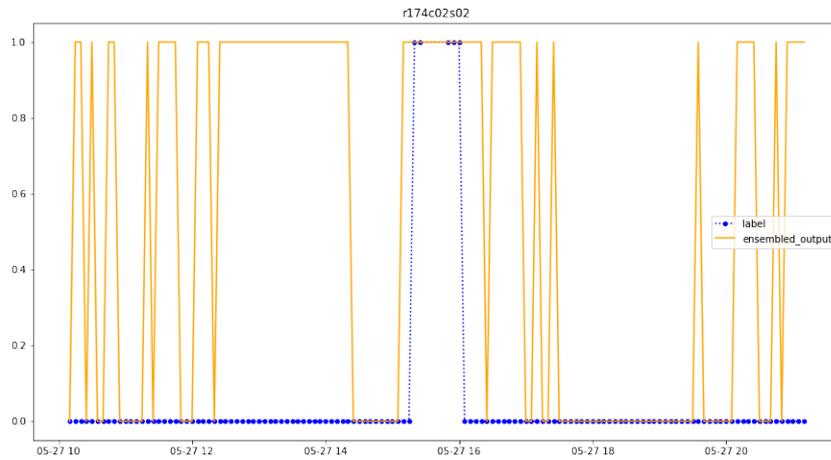


Figura 3.113: Output ensemble model anomalia 27 Maggio, alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	4.0	3590.0	13.0	16.0	3494.0	1.0	7118.0
r141c09s02	2.0	3607.0	17.0	5.0	3579.0	0.0	7210.0
r140c12s04	5.0	3663.0	40.0	29.0	3472.0	0.0	7209.0
r174c02s02	5.0	3706.0	15.0	13.0	3473.0	0.0	7212.0
r173c10s01	4.0	3552.0	23.0	16.0	3559.0	1.0	7155.0
r130c12s04	110.0	3560.0	30.0	9.0	3488.0	1.0	7198.0
r162c05s02	1321.0	2722.0	143.0	107.0	2854.0	51.0	7165.0
r129c12s02	2.0	3543.0	23.0	2.0	3553.0	0.0	7123.0
r183c05s01	4.0	3564.0	16.0	10.0	3611.0	0.0	7205.0
r129c12s03	2.0	3565.0	19.0	9.0	3486.0	0.0	7081.0
r141c09s01	1.0	3595.0	22.0	6.0	3582.0	1.0	7207.0
r135c09s03	1.0	3569.0	22.0	1.0	3550.0	1.0	7144.0
r162c05s03	1359.0	2560.0	152.0	124.0	3035.0	17.0	7165.0
r136c04s04	1.0	3555.0	10.0	5.0	3583.0	1.0	7155.0
r183c05s02	863.0	3179.0	40.0	10.0	3104.0	2.0	7198.0
r174c02s01	6.0	3659.0	16.0	4.0	3526.0	2.0	7213.0
r172c02s01	3.0	3666.0	21.0	12.0	3484.0	2.0	7188.0
r136c03s01	1.0	3586.0	15.0	11.0	3509.0	1.0	7123.0
r130c12s03	112.0	3569.0	31.0	24.0	3465.0	0.0	7201.0
r183c05s03	7.0	3552.0	17.0	9.0	3624.0	0.0	7209.0
mean	190.7	3478.1	34.2	21.1	3451.6	4.0	7173.9

Tabella 3.43: Risultati Maggio v1 con errore quadrato medio e alfa 0.9

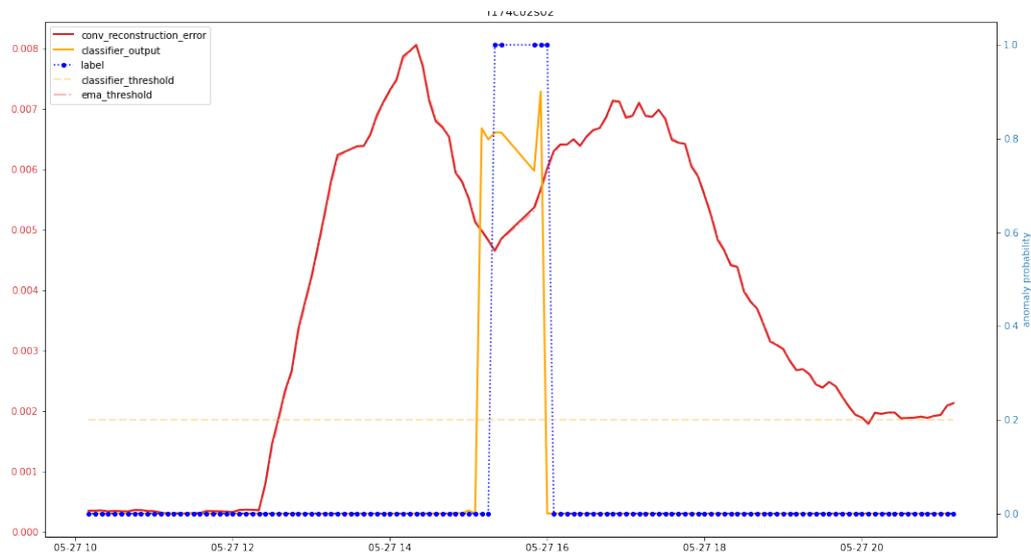


Figura 3.114: r174c02s02 anomalia 27 Maggio, alfa 0.9

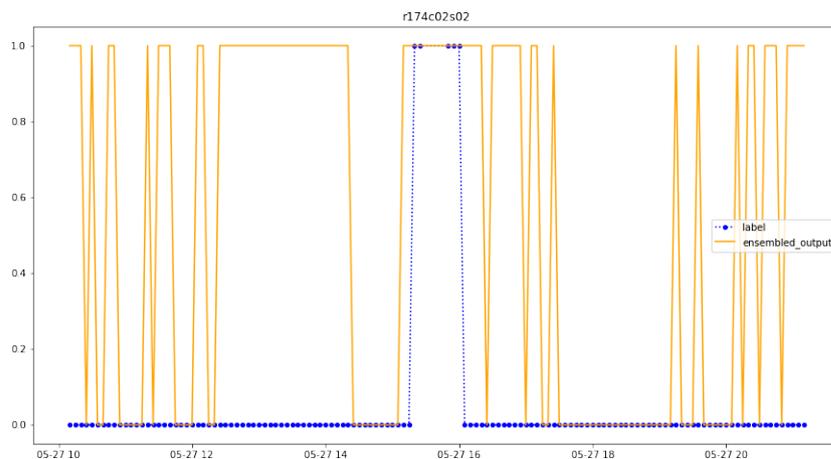


Figura 3.115: Output ensemble model anomalia 27 Maggio, alfa 0.9

Analisi Maggio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	3.0	3573.0	17.0	16.0	3507.0	2.0	7118.0
r141c09s02	2.0	3708.0	18.0	0.0	3482.0	0.0	7210.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r140c12s04	5.0	3720.0	42.0	15.0	3427.0	0.0	7209.0
r174c02s02	5.0	3716.0	16.0	11.0	3464.0	0.0	7212.0
r173c10s01	5.0	3571.0	24.0	15.0	3540.0	0.0	7155.0
r130c12s04	111.0	3579.0	39.0	4.0	3465.0	0.0	7198.0
r162c05s02	1329.0	2740.0	144.0	104.0	2840.0	43.0	7165.0
r129c12s02	2.0	3568.0	24.0	2.0	3527.0	0.0	7123.0
r183c05s01	4.0	3627.0	14.0	11.0	3549.0	0.0	7205.0
r129c12s03	2.0	3631.0	17.0	9.0	3422.0	0.0	7081.0
r141c09s01	2.0	3671.0	23.0	3.0	3508.0	0.0	7207.0
r135c09s03	1.0	3649.0	24.0	0.0	3469.0	1.0	7144.0
r162c05s03	1358.0	2702.0	155.0	102.0	2907.0	18.0	7165.0
r136c04s04	1.0	3574.0	18.0	3.0	3558.0	1.0	7155.0
r183c05s02	864.0	3224.0	35.0	11.0	3063.0	1.0	7198.0
r174c02s01	6.0	3665.0	17.0	1.0	3522.0	2.0	7213.0
r172c02s01	5.0	3738.0	22.0	14.0	3409.0	0.0	7188.0
r136c03s01	1.0	3670.0	16.0	5.0	3430.0	1.0	7123.0
r130c12s03	112.0	3659.0	39.0	25.0	3366.0	0.0	7201.0
r183c05s03	7.0	3644.0	13.0	11.0	3534.0	0.0	7209.0
mean	191.2	3531.4	35.9	18.1	3399.4	3.5	7173.9

Tabella 3.45: Risultati Maggio v1 con errore assoluto medio e alfa 0.3

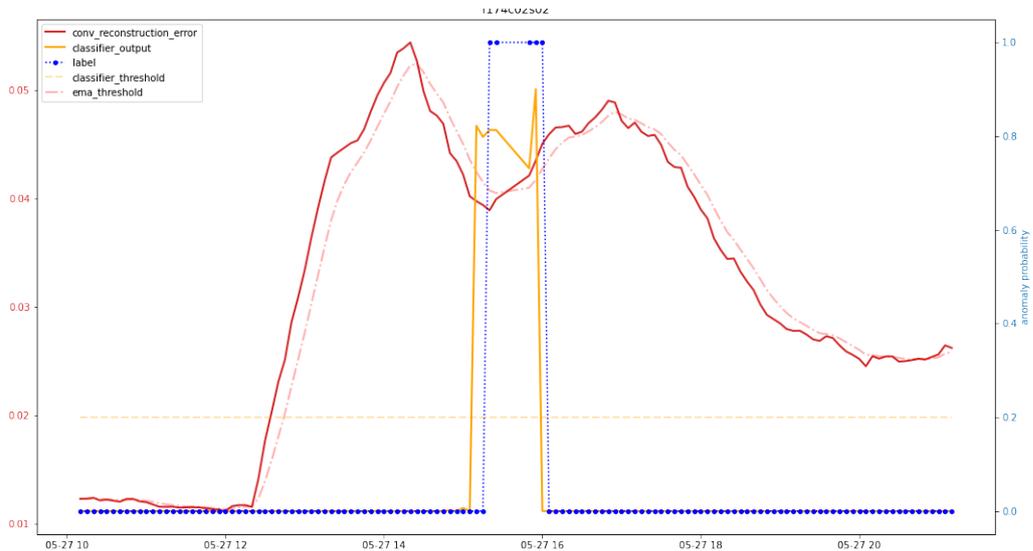


Figura 3.116: r174c02s02 anomalia 27 Maggio, alfa 0.3

Alfa	Anticipo medio
0.3	1 h 47 min
0.5	1 h 47 min
0.7	1 h 49 min
0.9	1 h 47 min

Tabella 3.44: Tabella anticipo medio Maggio v1 - errore quadrato

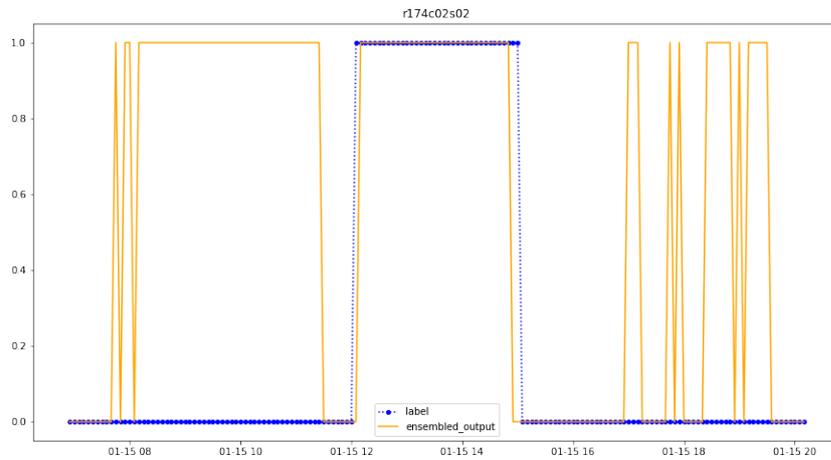


Figura 3.117: Output ensembled model anomalia 27 Maggio, alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	4.0	3598.0	14.0	16.0	3485.0	1.0	7118.0
r141c09s02	2.0	3673.0	18.0	2.0	3515.0	0.0	7210.0
r140c12s04	5.0	3735.0	43.0	12.0	3414.0	0.0	7209.0
r174c02s02	5.0	3714.0	15.0	9.0	3469.0	0.0	7212.0
r173c10s01	5.0	3567.0	23.0	15.0	3545.0	0.0	7155.0
r130c12s04	111.0	3553.0	38.0	5.0	3491.0	0.0	7198.0
r162c05s02	1329.0	2735.0	146.0	106.0	2842.0	43.0	7165.0
r129c12s02	2.0	3583.0	24.0	1.0	3513.0	0.0	7123.0
r183c05s01	4.0	3595.0	15.0	11.0	3580.0	0.0	7205.0
r129c12s03	2.0	3589.0	17.0	9.0	3464.0	0.0	7081.0
r141c09s01	2.0	3631.0	23.0	3.0	3548.0	0.0	7207.0
r135c09s03	1.0	3630.0	23.0	1.0	3488.0	1.0	7144.0
r162c05s03	1356.0	2706.0	154.0	107.0	2902.0	20.0	7165.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c04s04	1.0	3587.0	17.0	4.0	3545.0	1.0	7155.0
r183c05s02	864.0	3182.0	36.0	10.0	3105.0	1.0	7198.0
r174c02s01	6.0	3714.0	16.0	2.0	3473.0	2.0	7213.0
r172c02s01	5.0	3667.0	21.0	14.0	3481.0	0.0	7188.0
r136c03s01	1.0	3664.0	15.0	9.0	3433.0	1.0	7123.0
r130c12s03	112.0	3621.0	38.0	24.0	3406.0	0.0	7201.0
r183c05s03	7.0	3620.0	13.0	8.0	3561.0	0.0	7209.0
mean	191.2	3518.2	35.5	18.4	3413.0	3.5	7173.9

Tabella 3.46: Risultati Maggio v1 con errore assoluto medio e alfa 0.5

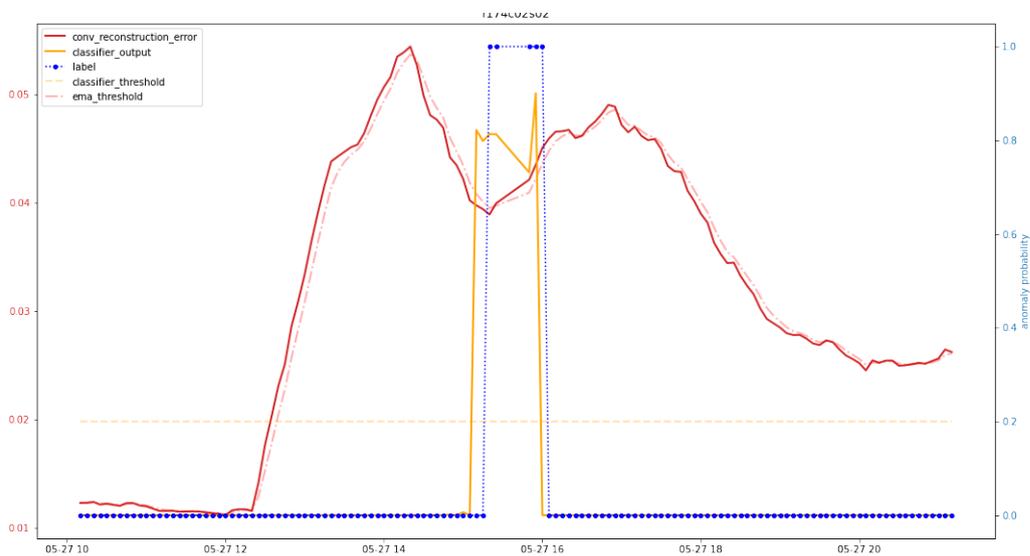


Figura 3.118: r174c02s02 anomalia 27 Maggio, alfa 0.5

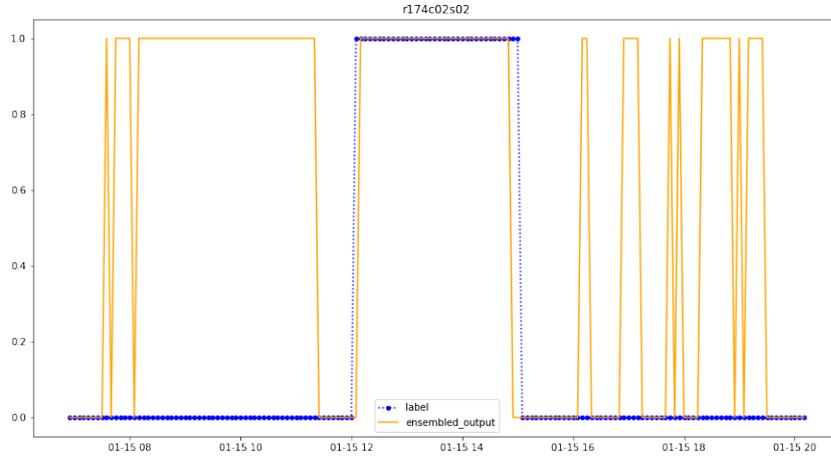


Figura 3.119: Output ensemble model anomalia 27 Maggio, alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	4.0	3591.0	13.0	17.0	3492.0	1.0	7118.0
r141c09s02	2.0	3661.0	16.0	3.0	3528.0	0.0	7210.0
r140c12s04	5.0	3728.0	42.0	12.0	3422.0	0.0	7209.0
r174c02s02	5.0	3703.0	15.0	11.0	3478.0	0.0	7212.0
r173c10s01	5.0	3576.0	23.0	15.0	3536.0	0.0	7155.0
r130c12s04	110.0	3559.0	38.0	4.0	3486.0	1.0	7198.0
r162c05s02	1330.0	2714.0	147.0	107.0	2863.0	42.0	7165.0
r129c12s02	2.0	3568.0	23.0	1.0	3529.0	0.0	7123.0
r183c05s01	4.0	3598.0	16.0	10.0	3577.0	0.0	7205.0
r129c12s03	2.0	3572.0	18.0	9.0	3480.0	0.0	7081.0
r141c09s01	2.0	3607.0	23.0	3.0	3572.0	0.0	7207.0
r135c09s03	1.0	3598.0	23.0	1.0	3520.0	1.0	7144.0
r162c05s03	1356.0	2678.0	153.0	109.0	2930.0	20.0	7165.0
r136c04s04	1.0	3581.0	16.0	4.0	3552.0	1.0	7155.0
r183c05s02	864.0	3166.0	36.0	10.0	3121.0	1.0	7198.0
r174c02s01	6.0	3729.0	16.0	3.0	3457.0	2.0	7213.0
r172c02s01	3.0	3632.0	21.0	14.0	3516.0	2.0	7188.0
r136c03s01	1.0	3622.0	14.0	9.0	3476.0	1.0	7123.0
r130c12s03	112.0	3622.0	37.0	24.0	3406.0	0.0	7201.0
r183c05s03	7.0	3588.0	13.0	7.0	3594.0	0.0	7209.0
mean	191.1	3504.7	35.1	18.6	3426.8	3.6	7173.9

Continua nella prossima pagina

tp	tn	fp_pre	fp_post	fp_random	fn	records
----	----	--------	---------	-----------	----	---------

Tabella 3.47: Risultati Maggio v1 con errore assoluto medio e alfa 0.7

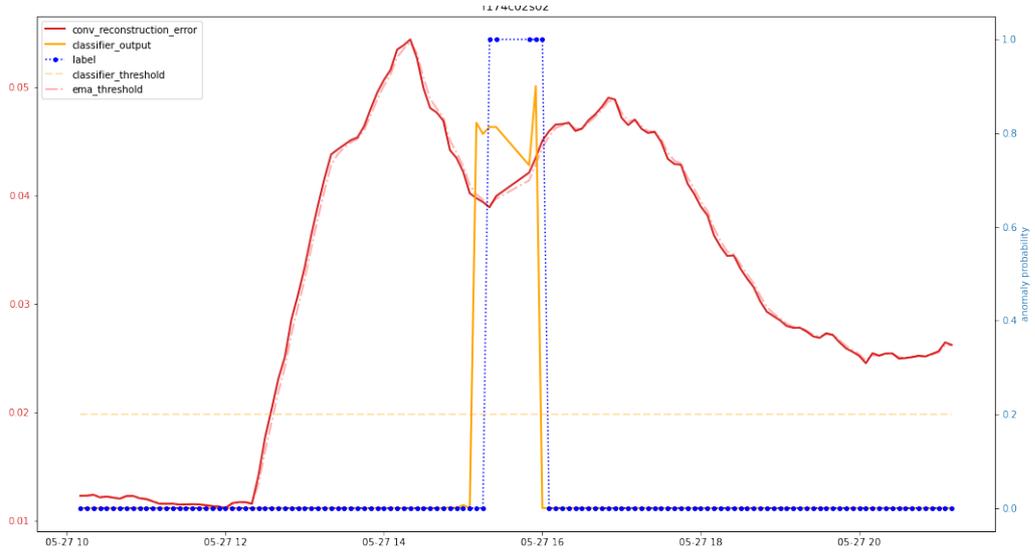


Figura 3.120: r174c02s02 anomalia 27 Maggio, alfa 0.7

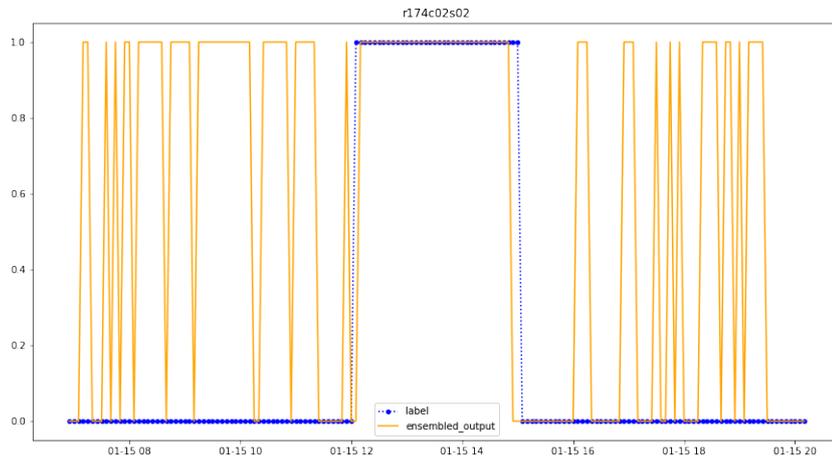


Figura 3.121: Output ensemble model anomalia 27 Maggio, alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	4.0	3562.0	12.0	17.0	3522.0	1.0	7118.0
r141c09s02	1.0	3634.0	15.0	5.0	3554.0	1.0	7210.0
r140c12s04	5.0	3714.0	40.0	17.0	3433.0	0.0	7209.0
r174c02s02	5.0	3664.0	15.0	11.0	3517.0	0.0	7212.0
r173c10s01	4.0	3585.0	22.0	15.0	3528.0	1.0	7155.0
r130c12s04	110.0	3551.0	37.0	4.0	3495.0	1.0	7198.0
r162c05s02	1328.0	2751.0	147.0	107.0	2826.0	44.0	7165.0
r129c12s02	2.0	3574.0	22.0	2.0	3523.0	0.0	7123.0
r183c05s01	4.0	3552.0	17.0	10.0	3622.0	0.0	7205.0
r129c12s03	2.0	3552.0	18.0	9.0	3500.0	0.0	7081.0
r141c09s01	1.0	3574.0	22.0	5.0	3604.0	1.0	7207.0
r135c09s03	1.0	3598.0	23.0	2.0	3519.0	1.0	7144.0
r162c05s03	1356.0	2647.0	150.0	112.0	2962.0	20.0	7165.0
r136c04s04	1.0	3592.0	14.0	4.0	3543.0	1.0	7155.0
r183c05s02	864.0	3138.0	34.0	10.0	3151.0	1.0	7198.0
r174c02s01	6.0	3716.0	16.0	4.0	3469.0	2.0	7213.0
r172c02s01	3.0	3630.0	22.0	12.0	3519.0	2.0	7188.0
r136c03s01	1.0	3601.0	16.0	10.0	3494.0	1.0	7123.0
r130c12s03	112.0	3625.0	37.0	24.0	3403.0	0.0	7201.0
r183c05s03	7.0	3589.0	14.0	7.0	3592.0	0.0	7209.0
mean	190.8	3492.4	34.6	19.4	3438.8	3.9	7173.9

Tabella 3.48: Risultati Maggio v1 con errore assoluto medio e alfa 0.9

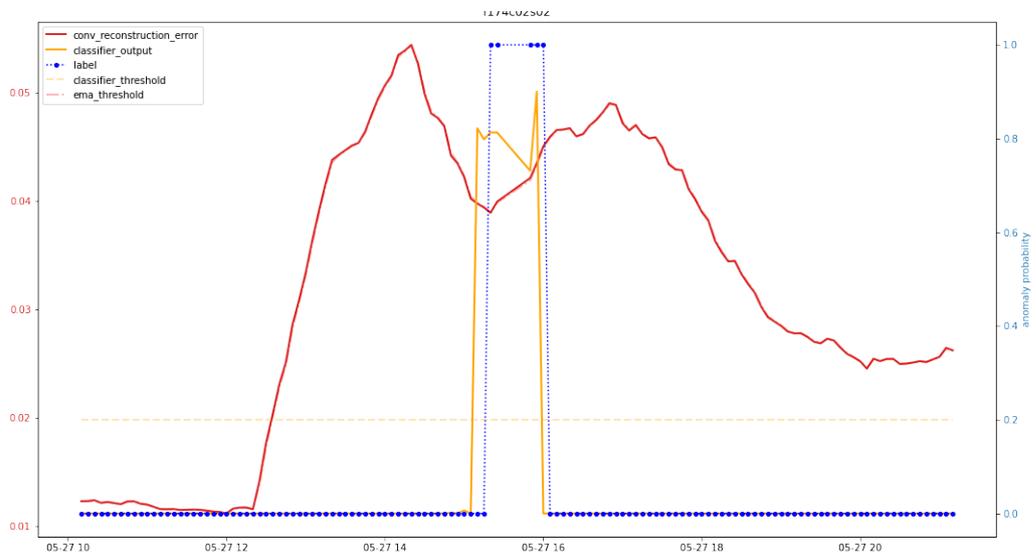


Figura 3.122: r174c02s02 anomalia 27 Maggio, alfa 0.9

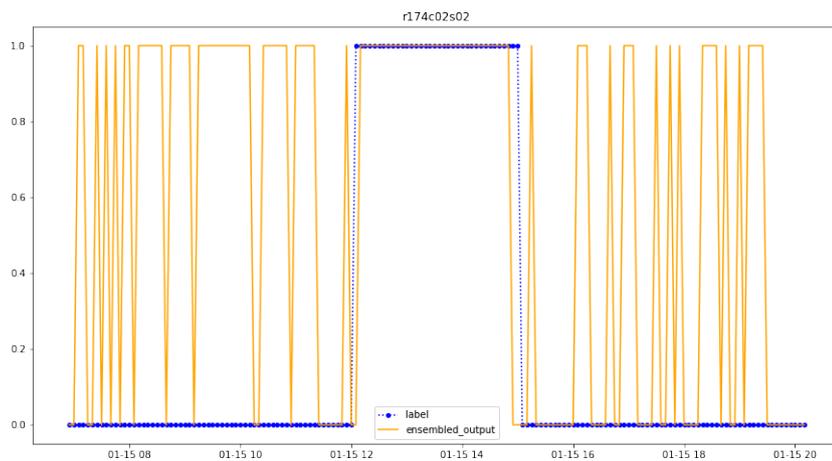


Figura 3.123: Output ensemble model anomalia 27 Maggio, alfa 0.9

Il trend è definitivamente confermato, c'è una presenza preponderante di falsi positivi non giustificati da anomalie, per cui nella sezione successiva analizzerò quello che succede effettivamente in quel periodo.

Si può anche notare come in questi esperimenti la variazione di alfa e della metrica di errore assoluto e errore quadrato non influenzi particolarmente sia il numero di falsi positivi sia l'anticipo medio.

Alfa	Anticipo medio
0.3	1 h 46 min
0.5	1 h 47 min
0.7	1 h 47 min
0.9	1 h 45 min

Tabella 3.49: Tabella anticipo medio Maggio v1 - errore assoluto

3.3.3 Analisi random false positives

Come si può vedere nelle tabelle si hanno ottimi risultati in termini di veri positivi, falsi positivi pre-anomalia, falsi positivi post anomalia, e falsi negativi. Occorre quindi capire cosa causa la presenza preponderante di falsi positivi che ho denominato come random poiché non associabili a un'anomalia.

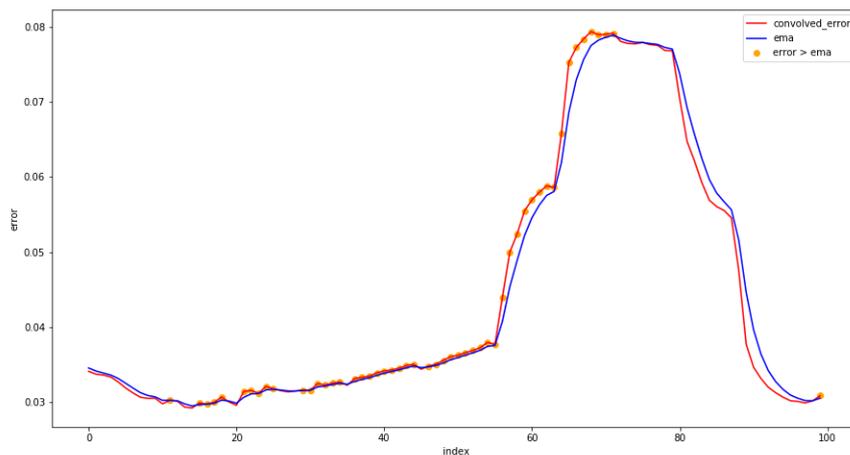


Figura 3.124: Esempio di random_fp di r129c12s02

Prendendo come esempio l'immagine 3.124, nella quale viene presentato un periodo non anomalo, si può notare un grande numero di casi in cui EMA è maggiore dell'errore, come indicato nelle tabelle mostrate precedentemente.

Si può notare sempre in figura 3.124 come la presenza di questi falsi positivi sia dovuta al trend di crescita dell'errore, in realtà questo fenomeno è molto semplice da spiegare.

Se a EMA vengono presentati valori con un trend di crescita, EMA incrementerà il proprio valore, ma a meno che alfa sia pari a uno, il valore di

EMA sarà minore di quello presentato. Inoltre in molti falsi positivi si può osservare come in realtà EMA sia di poco inferiore all'errore.

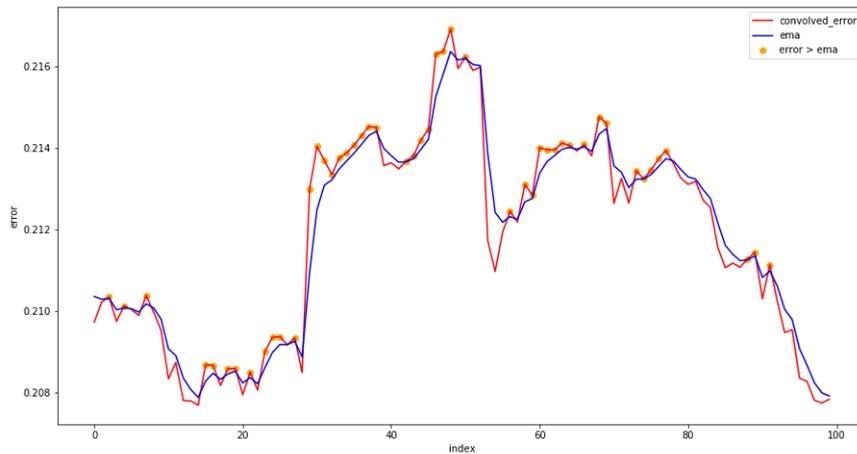


Figura 3.125: Esempio di random_fp di r140c12s02

Per evidenziare il problema ho preso lo stesso periodo, ovviamente non anomalo, in un altro nodo. In figura 3.125 il grafico sembra molto più variabile ma in realtà la variazione d'errore è molto piccola e si può vedere che quando l'errore cresce la presenza di falsi positivi incrementa.

La soluzione sarebbe rendere la condizione di anomalia di AE-semisupervised più stringente, quindi non superare EMA ma superare EMA più un valore che possa ridurre il tasso dei falsi positivi. Nel capitolo successivo esplorerò due soluzioni su quale valore utilizzare.

Riporto alcune tabelle riassuntive (Tabella 3.50, Tabella 3.51) sulle performance del modello v2 confrontate con v1. Ho preso come riferimento alfa uguale a 0.5 al fine di rendere le tabelle il più compatte possibile, siccome nella v1 errore assoluto e quadrato non hanno molte differenze tra loro, ho scelto come riferimento di v1 l'errore quadrato.

Metrica	v1 - MSE	v2 - MSE
tp	21.6	19.7
tn	2859.7	5668.9
fp_pre	38.9	2.0
fp_post	18.3	4.7
fp_random	2762.8	2.9
fn	0.9	2.8

Tabella 3.50: Confronto v1 - v2-Mean, alfa 0.5, Gennaio

Metrica	v1 - MSE	v2 - MSE
tp	190.8	186.4
tn	3495.2	6932.1
fp_pre	35.1	10.0
fp_post	20.9	9.6
fp_random	3433.6	32.6
fn	4.0	8.3

Tabella 3.51: Confronto v1 - v2-Mean, alfa 0.5, Maggio

3.3.4 Analisi modello ensembled v2 - Mean

La prima considerazione da fare sul valore da trovare per ridurre il numero di falsi positivi è che non può essere un valore identico per ogni nodo, ma differente e coerente coi valori presenti in quel dataset, in modo che sia un valore significativo. Premetto che questi risultati sono frutto di test del tutto empirici, fatti col fine di esplorare possibili soluzioni al problema dell'alto tasso di falsi positivi.

Ho scelto di utilizzare la media dei "mean_error" di tutto il dataset come valore da aggiungere all'EMA poiché dovrebbe avere un valore coerente con l'andamento generale dell'errore.

Nelle pagine seguenti riporto i risultati dell'esperimento al fine di compararli con i risultati precedenti e l'output del modello ensembled delle stesse anomalie mostrate nella modello v1.

Non mostrerò tutte le tabelle riguardanti l'utilizzo dell'errore assoluto perché sono tutte identiche, le performance non variano al variare di alfa, il motivo sarà spiegato nella sezione di confronto con altre tecniche.

Analisi Gennaio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6897.0	3.0	18.0	25.0	0.0	6958.0
r130c12s04	6.0	6985.0	0.0	3.0	1.0	0.0	6995.0
r162c05s02	64.0	6844.0	4.0	3.0	10.0	6.0	6931.0
r129c12s02	10.0	6920.0	10.0	4.0	4.0	2.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	12.0	6941.0	9.0	0.0	0.0	2.0	6964.0
r141c09s01	6.0	6948.0	11.0	6.0	14.0	6.0	6991.0
r135c09s03	22.0	6869.0	10.0	16.0	2.0	1.0	6899.0
r162c05s03	64.0	6840.0	1.0	3.0	9.0	9.0	6926.0
r136c04s04	18.0	6891.0	13.0	13.0	18.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2608.0	0.0	2.0	2.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	7.0	6904.0	6.0	4.0	0.0	1.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	20.2	5664.4	3.7	5.1	5.4	2.4	5698.2

Tabella 3.52: Risultati Gennaio v2 - Mean con errore quadrato medio e alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6908.0	1.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6931.0	3.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6948.0	2.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6970.0	3.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6910.0	11.0	13.0	1.0	3.0	6936.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6910.0	1.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5668.9	2.0	4.7	2.9	2.8	5698.2

Tabella 3.53: Risultati Gennaio v2 - Mean con errore quadrato medio e alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6909.0	0.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6911.0	11.0	13.0	0.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6911.0	0.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.5	1.5	4.7	2.9	2.8	5698.2

Tabella 3.54: Risultati Gennaio v2 - Mean con errore quadrato medio e alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6909.0	0.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6911.0	11.0	13.0	0.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6911.0	0.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.5	1.5	4.7	2.9	2.8	5698.2

Tabella 3.55: Risultati Gennaio v2 - Mean con errore quadrato medio e alfa 0.9

Alfa	Anticipo medio
0.3	0 h 33 min
0.5	0 h 33 min
0.7	0 h 33 min
0.9	0 h 33 min

Tabella 3.56: Tabella anticipo medio Gennaio v2 Mean - errore quadrato

Nonostante il numero di falsi positivi pre-anomalia (fp_pre) sia variabile al variare di alfa (seppur di poco), l'anticipo medio non cambia perché molto probabilmente non si presentano sequenze di 3 falsi positivi pre-anomalia vicini tra loro anche in altre anomalie.

Analisi Gennaio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6909.0	0.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6911.0	11.0	13.0	0.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6911.0	0.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.5	1.5	4.7	2.9	2.8	5698.2

Tabella 3.57: Risultati Gennaio v2 - Mean con errore assoluto medio e alfa 0.3

Alfa	Anticipo medio
0.3	0 h 33 min
0.5	0 h 33 min
0.7	0 h 33 min
0.9	0 h 33 min

Tabella 3.58: Tabella anticipo medio Gennaio v2 Mean - errore assoluto

Analisi Maggio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7102.0	0.0	0.0	11.0	3.0	7118.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r141c09s02	1.0	7206.0	0.0	0.0	2.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7187.0	2.0	0.0	18.0	1.0	7212.0
r173c10s01	2.0	7145.0	0.0	4.0	1.0	3.0	7155.0
r130c12s04	110.0	7066.0	5.0	1.0	15.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7195.0	1.0	0.0	5.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7140.0	0.0	0.0	2.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7188.0	9.0	1.0	7.0	2.0	7213.0
r172c02s01	0.0	7176.0	0.0	0.0	7.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7043.0	1.0	24.0	21.0	1.0	7201.0
r183c05s03	4.0	7190.0	0.0	0.0	12.0	3.0	7209.0
mean	186.4	6927.6	10.4	9.6	36.7	8.3	7173.9

Tabella 3.59: Risultati Maggio v2 - Mean con errore quadrato medio e alfa 0.3

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7207.0	0.0	0.0	1.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7198.0	1.0	0.0	2.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7192.0	0.0	0.0	10.0	3.0	7209.0
mean	186.4	6932.1	10.0	9.6	32.6	8.3	7173.9

Tabella 3.60: Risultati Maggio v2 - Mean con errore quadrato medio e alfa 0.5

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6932.2	10.0	9.6	32.5	8.3	7173.9

Tabella 3.61: Risultati Maggio v2 - Mean con errore quadrato medio e alfa 0.7

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6932.2	10.0	9.6	32.5	8.3	7173.9

Tabella 3.62: Risultati Maggio v2 - Mean con errore quadrato medio e alfa 0.9

Alfa	Anticipo medio
0.3	0 h 58 min
0.5	0 h 57 min
0.7	0 h 57 min
0.9	0 h 57 min

Tabella 3.63: Tabella anticipo medio Maggio v2 Mean - errore quadrato

Analisi Maggio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6932.2	10.0	9.6	32.5	8.3	7173.9

Tabella 3.64: Risultati Maggio v2 - Mean con errore assoluto medio e alfa 0.3

Alfa	Anticipo medio
0.3	0 h 58 min
0.5	0 h 58 min
0.7	0 h 58 min
0.9	0 h 58 min

Tabella 3.65: Tabella anticipo medio Maggio v2 Mean - errore assoluto

Come detto in precedenza, utilizzando MAE si hanno risultati (e quindi anche performance) identiche al variare di alfa, ciò giustifica l'uguaglianza in termini di anticipo al variare di alfa in tabella 3.65.

I miglioramenti rispetto alla versione 1 sono netti, soprattutto per quanto riguarda i falsi positivi denominati random.

Questa soluzione ha però un problema, se si volesse applicare questo modello ensemble a un sistema online, di cui quindi non si conosce l'intero dataset a priori ma i record arrivano incrementalmente, questo approccio non è direttamente utilizzabile.

Nella sezione successiva quindi introduco una modifica che consentirebbe al modello ensemble di lavorare in un contesto online.

3.3.5 Analisi modello ensemble v2 - Window Mean

Dato il miglioramento che si è ottenuto usando una media dell'errore medio su tutto il periodo, ho deciso di mantenere il concetto di media, ma quello che occorre cambiare è il set di dati su cui è calcolata, in particolare bisogna che sia utilizzabile in un contesto in cui i dati arrivano in modo incrementale.

Si è quindi passati da calcolare una media classica a una media a finestra, in particolare si usa come valore la media dell'errore medio degli ultimi 5 record. In questo modo anche in un contesto dove non sono presenti tutti i record si può effettuare questo tipo di correzione per ridurre il numero di falsi positivi.

Di seguito riporto i risultati ottenuti usando questo tipo di approccio.

Analisi Gennaio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6908.0	0.0	7.0	1.0	0.0	6997.0
r174c02s02	40.0	6918.0	4.0	6.0	2.0	5.0	6975.0
r173c10s01	15.0	6873.0	11.0	18.0	41.0	0.0	6958.0
r130c12s04	6.0	6938.0	7.0	3.0	41.0	0.0	6995.0
r162c05s02	64.0	6808.0	2.0	2.0	49.0	6.0	6931.0
r129c12s02	7.0	6931.0	1.0	0.0	6.0	5.0	6950.0
r183c05s01	6.0	1509.0	0.0	2.0	9.0	0.0	1518.0
r129c12s03	9.0	6943.0	2.0	0.0	5.0	5.0	6964.0
r141c09s01	6.0	6959.0	0.0	6.0	14.0	6.0	6991.0
r135c09s03	22.0	6832.0	23.0	16.0	26.0	1.0	6899.0
r162c05s03	64.0	6792.0	1.0	2.0	58.0	9.0	6926.0
r136c04s04	18.0	6880.0	19.0	13.0	23.0	3.0	6936.0
r183c05s02	6.0	2622.0	0.0	1.0	3.0	0.0	2632.0
r174c02s01	1.0	2606.0	1.0	2.0	3.0	2.0	2615.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r172c02s01	3.0	1520.0	0.0	5.0	9.0	0.0	1529.0
r136c03s01	5.0	6887.0	10.0	3.0	14.0	3.0	6922.0
r130c12s03	10.0	6951.0	0.0	0.0	12.0	5.0	6978.0
r183c05s03	6.0	2611.0	0.0	1.0	7.0	0.0	2625.0
mean	19.7	5652.6	4.3	4.7	17.0	2.8	5698.2

Tabella 3.66: Risultati Gennaio v2 - Window Mean con errore quadrato medio e alfa 0.3

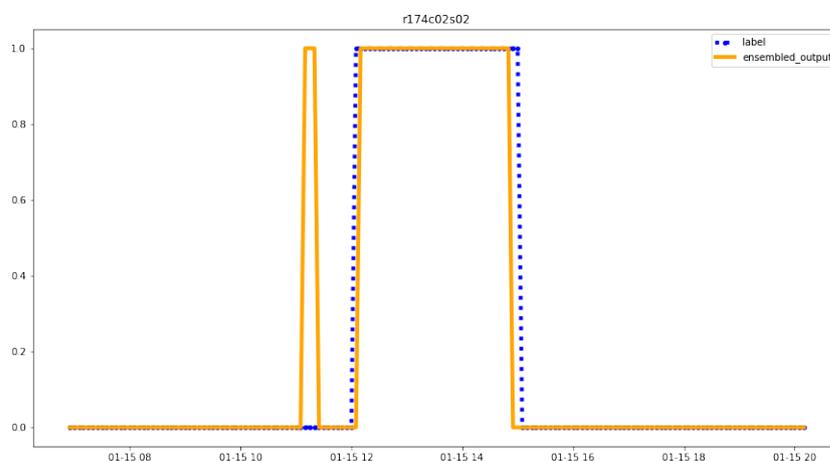


Figura 3.126: Output ensembled model anomalia 15 gennaio, alfa 0.3, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6896.0	3.0	18.0	26.0	0.0	6958.0
r130c12s04	6.0	6967.0	3.0	3.0	16.0	0.0	6995.0
r162c05s02	64.0	6844.0	2.0	2.0	13.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r141c09s01	6.0	6962.0	0.0	6.0	11.0	6.0	6991.0
r135c09s03	22.0	6852.0	18.0	16.0	11.0	1.0	6899.0
r162c05s03	64.0	6831.0	1.0	2.0	19.0	9.0	6926.0
r136c04s04	18.0	6903.0	12.0	13.0	7.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6900.0	8.0	3.0	3.0	3.0	6922.0
r130c12s03	10.0	6961.0	0.0	0.0	2.0	5.0	6978.0
r183c05s03	6.0	2612.0	0.0	1.0	6.0	0.0	2625.0
mean	19.7	5664.1	2.7	4.7	7.1	2.8	5698.2

Tabella 3.67: Risultati Gennaio v2 - Window Mean con errore quadrato medio e alfa 0.5

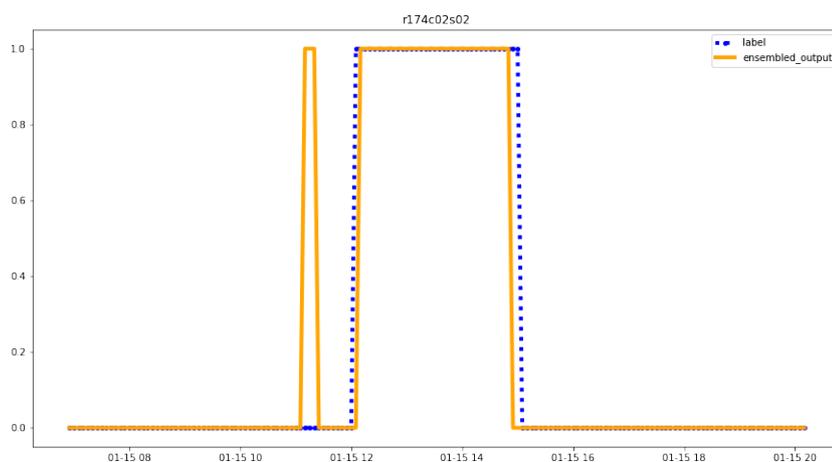


Figura 3.127: Output ensembled model anomalia 15 gennaio, alfa 0.5, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r173c10s01	15.0	6906.0	1.0	18.0	18.0	0.0	6958.0
r130c12s04	6.0	6985.0	0.0	3.0	1.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6970.0	0.0	6.0	3.0	6.0	6991.0
r135c09s03	22.0	6864.0	15.0	16.0	2.0	1.0	6899.0
r162c05s03	64.0	6839.0	1.0	2.0	11.0	9.0	6926.0
r136c04s04	18.0	6908.0	11.0	13.0	3.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6910.0	1.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5668.5	1.9	4.7	3.5	2.8	5698.2

Tabella 3.68: Risultati Gennaio v2 - Window Mean con errore quadrato medio e alfa 0.7

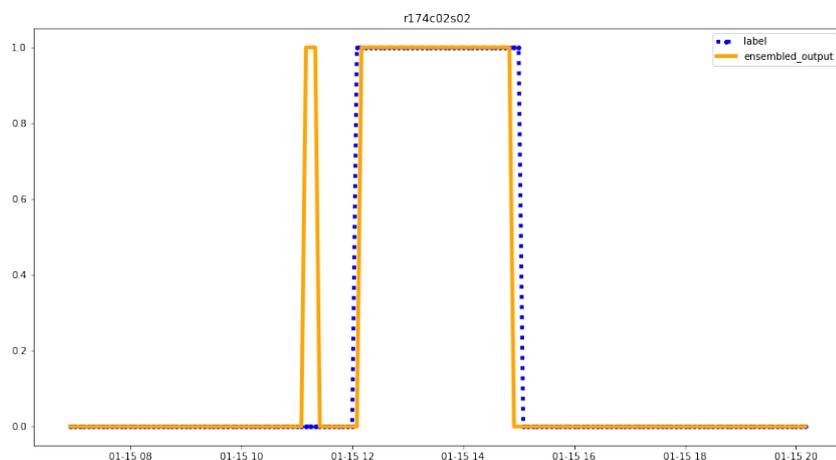


Figura 3.128: Output ensembled model anomalia 15 gennaio, alfa 0.7, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6909.0	0.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6911.0	11.0	13.0	0.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6911.0	0.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.5	1.5	4.7	2.9	2.8	5698.2

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
--	----	----	--------	---------	-----------	----	---------

Tabella 3.69: Risultati Gennaio v2 - Window Mean con errore quadrato medio e alfa 0.9

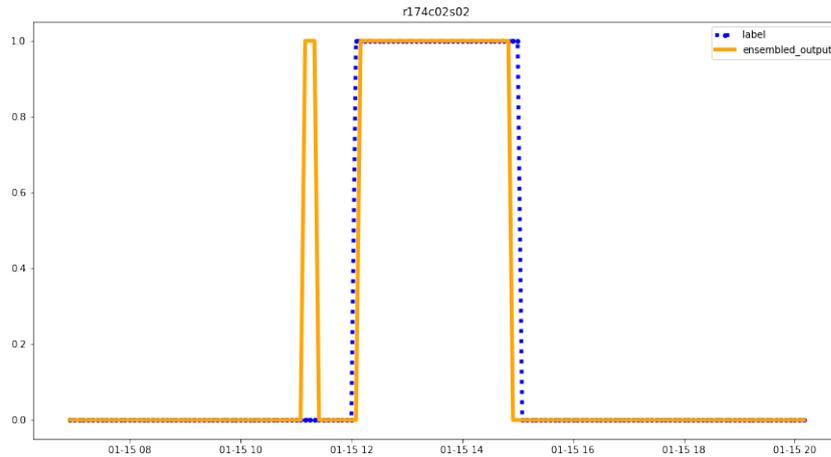


Figura 3.129: Output ensemble model anomalia 15 gennaio, alfa 0.9, v2-window Mean

Alfa	Anticipo medio
0.3	0 h 42 min
0.5	0 h 33 min
0.7	0 h 33 min
0.9	0 h 33 min

Tabella 3.70: Tabella anticipo medio Gennaio v2 Window Mean - errore quadrato

Come si può vedere in tabella 3.70, con alfa maggiore di 0.5 l'anticipo medio non cambia perché nonostante in media calino al calare di alfa, rimangono presenti quei falsi positivi che consentono di effettuare la predizione in anticipo dell'anomalia.

Analisi Gennaio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6909.0	0.0	18.0	16.0	0.0	6958.0
r130c12s04	6.0	6986.0	0.0	3.0	0.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6841.0	1.0	2.0	9.0	9.0	6926.0
r136c04s04	18.0	6911.0	11.0	13.0	0.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6911.0	0.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.5	1.5	4.7	2.9	2.8	5698.2

Tabella 3.71: Risultati Gennaio v2 - Window Mean con errore assoluto medio e alfa 0.3

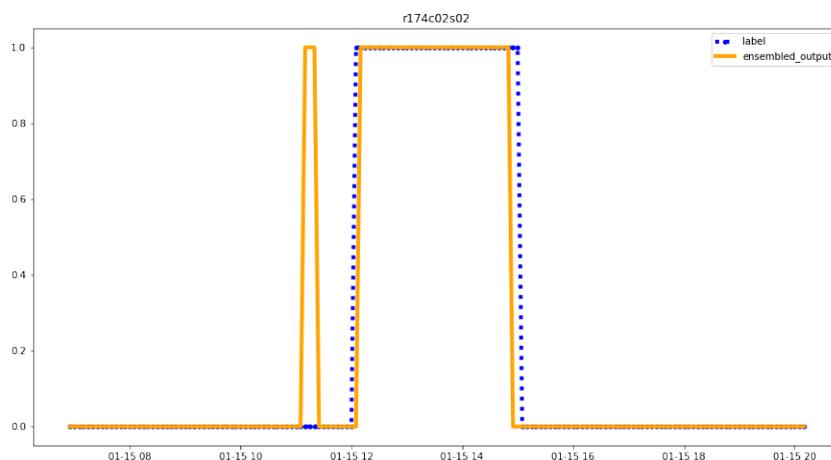


Figura 3.130: Output ensembled model anomalia 15 gennaio, alfa 0.3, v2-window Mean, MAE

Analisi Maggio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7088.0	0.0	0.0	25.0	3.0	7118.0
r141c09s02	1.0	7194.0	0.0	0.0	14.0	1.0	7210.0
r140c12s04	2.0	7200.0	0.0	1.0	3.0	3.0	7209.0
r174c02s02	4.0	7117.0	2.0	0.0	88.0	1.0	7212.0
r173c10s01	2.0	7117.0	0.0	4.0	29.0	3.0	7155.0
r130c12s04	110.0	7041.0	5.0	1.0	40.0	1.0	7198.0
r162c05s02	1278.0	5410.0	91.0	68.0	248.0	94.0	7165.0
r129c12s02	2.0	7065.0	2.0	0.0	54.0	0.0	7123.0
r183c05s01	1.0	7190.0	1.0	0.0	10.0	3.0	7205.0
r129c12s03	2.0	7047.0	7.0	0.0	25.0	0.0	7081.0
r141c09s01	1.0	7158.0	0.0	0.0	47.0	1.0	7207.0
r135c09s03	0.0	7109.0	0.0	0.0	33.0	2.0	7144.0
r162c05s03	1343.0	5255.0	96.0	93.0	422.0	33.0	7165.0
r136c04s04	1.0	7100.0	2.0	0.0	51.0	1.0	7155.0
r183c05s02	857.0	6329.0	0.0	1.0	3.0	8.0	7198.0
r174c02s01	6.0	7144.0	3.0	1.0	57.0	2.0	7213.0
r172c02s01	0.0	7118.0	0.0	0.0	65.0	5.0	7188.0
r136c03s01	1.0	7109.0	1.0	0.0	11.0	1.0	7123.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r130c12s03	111.0	7044.0	1.0	24.0	20.0	1.0	7201.0
r183c05s03	4.0	7185.0	0.0	0.0	17.0	3.0	7209.0
mean	186.4	6901.0	10.6	9.7	63.1	8.3	7173.9

Tabella 3.72: Risultati Maggio v2 - Window Mean con errore quadrato medio e alfa 0.3

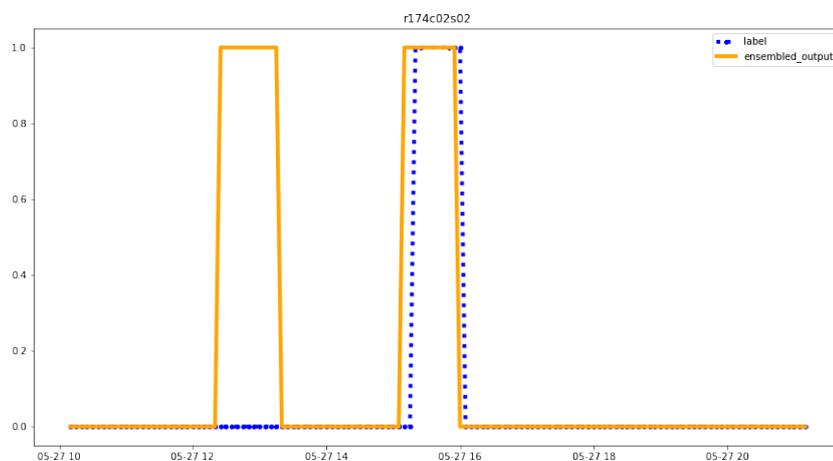


Figura 3.131: Output ensembled model anomalia 27 Maggio, alfa 0.3, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7106.0	0.0	0.0	7.0	3.0	7118.0
r141c09s02	1.0	7207.0	0.0	0.0	1.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7181.0	2.0	0.0	24.0	1.0	7212.0
r173c10s01	2.0	7143.0	0.0	4.0	3.0	3.0	7155.0
r130c12s04	110.0	7065.0	5.0	1.0	16.0	1.0	7198.0
r162c05s02	1278.0	5421.0	91.0	68.0	237.0	94.0	7165.0
r129c12s02	2.0	7103.0	1.0	0.0	17.0	0.0	7123.0
r183c05s01	1.0	7197.0	1.0	0.0	3.0	3.0	7205.0
r129c12s03	2.0	7073.0	2.0	0.0	4.0	0.0	7081.0
r141c09s01	1.0	7190.0	0.0	0.0	15.0	1.0	7207.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r135c09s03	0.0	7136.0	0.0	0.0	6.0	2.0	7144.0
r162c05s03	1343.0	5274.0	96.0	92.0	404.0	33.0	7165.0
r136c04s04	1.0	7140.0	2.0	0.0	11.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7171.0	3.0	1.0	30.0	2.0	7213.0
r172c02s01	0.0	7162.0	0.0	0.0	21.0	5.0	7188.0
r136c03s01	1.0	7118.0	1.0	0.0	2.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7192.0	0.0	0.0	10.0	3.0	7209.0
mean	186.4	6923.6	10.2	9.6	40.8	8.3	7173.9

Tabella 3.73: Risultati Maggio v2 - Window Mean con errore quadrato medio e alfa 0.5

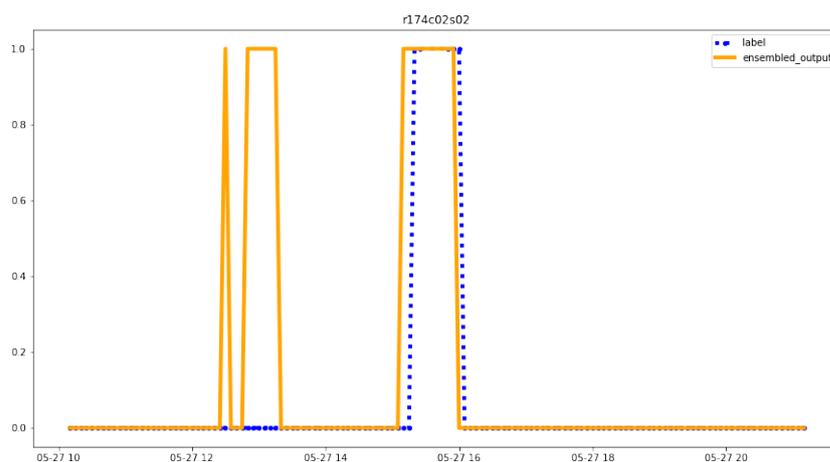


Figura 3.132: Output ensembled model anomalia 27 Maggio, alfa 0.5, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7111.0	0.0	0.0	2.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7200.0	2.0	0.0	5.0	1.0	7212.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r173c10s01	2.0	7145.0	0.0	4.0	1.0	3.0	7155.0
r130c12s04	110.0	7080.0	5.0	1.0	1.0	1.0	7198.0
r162c05s02	1278.0	5424.0	91.0	68.0	234.0	94.0	7165.0
r129c12s02	2.0	7116.0	1.0	0.0	4.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7077.0	1.0	0.0	1.0	0.0	7081.0
r141c09s01	1.0	7200.0	0.0	0.0	5.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7150.0	2.0	0.0	1.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7185.0	3.0	1.0	16.0	2.0	7213.0
r172c02s01	0.0	7177.0	0.0	0.0	6.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6930.1	10.2	9.6	34.4	8.3	7173.9

Tabella 3.74: Risultati Maggio v2 - Window Mean con errore quadrato medio e alfa 0.7

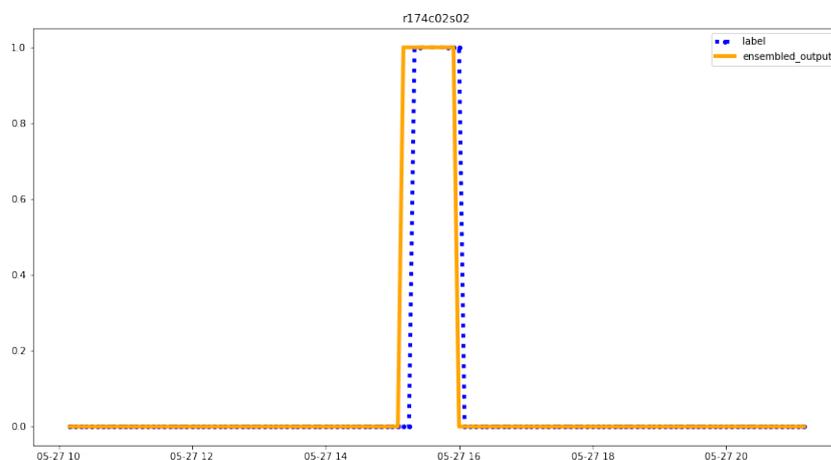


Figura 3.133: Output ensembled model anomalia 27 Maggio, alfa 0.7, v2-window Mean

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7203.0	0.0	0.0	2.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
mean	186.4	6932.2	10.0	9.6	32.5	8.3	7173.9

Tabella 3.75: Risultati Maggio v2 - Window Mean con errore quadrato medio e alfa 0.9

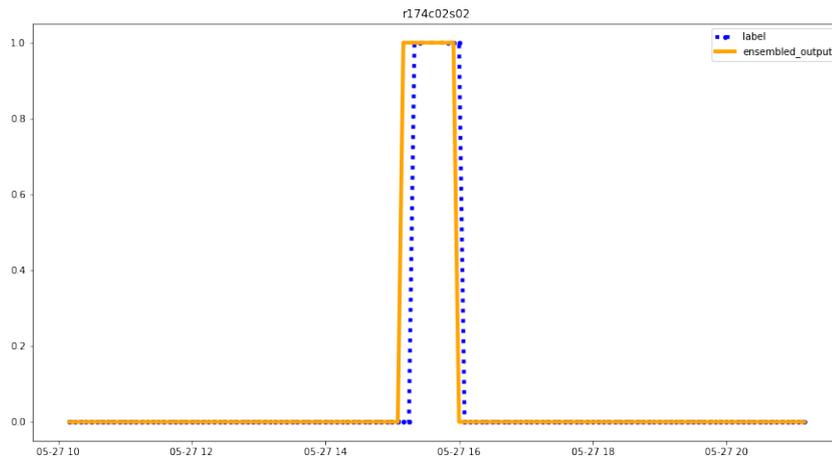


Figura 3.134: Output ensemble model anomalia 27 Maggio, alfa 0.9, v2-window Mean

Alfa	Anticipo medio
0.3	0 h 58 min
0.5	0 h 57 min
0.7	0 h 57 min
0.9	0 h 57 min

Tabella 3.76: Tabella anticipo medio Maggio v2 Window Mean - errore quadrato

Analisi Maggio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7204.0	0.0	0.0	1.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7183.0	0.0	0.0	0.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6932.2	10.0	9.6	32.5	8.3	7173.9

Tabella 3.77: Risultati Maggio v2 - Window Mean con errore assoluto medio e alfa 0.3

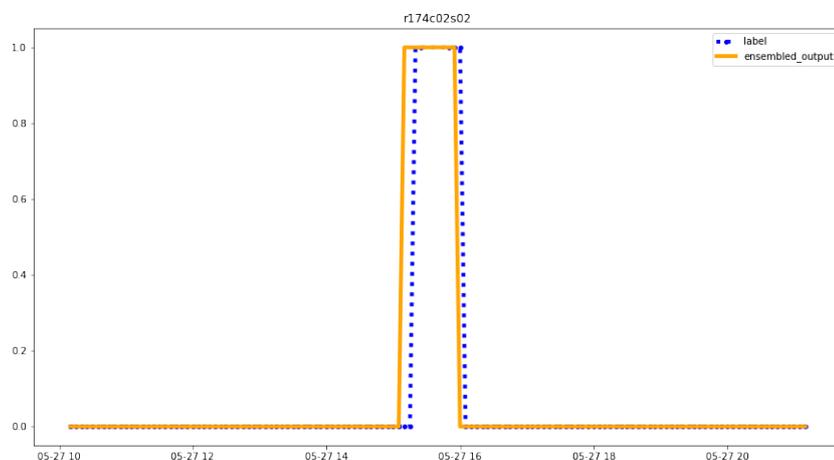


Figura 3.135: Output ensembled model anomalia 27 Maggio, alfa 0.3, v2-window Mean

I risultati con MSE e alfa pari a 0.9 sono molto simili (anche in termini di grafici presentati) ai risultati ottenuti con MAE.

3.3.6 Analisi modello ensembled v2 - WindowMean&sum

Questo esperimento è nato in modo del tutto casuale ma ha portato alla luce il problema di utilizzare un "valore" per regolarizzare il numero di falsi positivi random.

In particolare come si potrà vedere con dati alla mano nella sezione seguente, AE-semisupervised non si attiva quasi mai nelle versioni presentate fin ora poiché il divario da errore e EMA + window mean è troppo alto.

In questo esperimento si è deciso di utilizzare la somma dell'errore medio degli ultimi 24 record come valore di errore da considerare per la rilevazione e come metodo di detection per AE-semisupervised sempre EMA + window mean.

Analisi Gennaio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	6847.0	5.0	12.0	51.0	0.0	6924.0
r140c12s04	81.0	6600.0	14.0	8.0	294.0	0.0	6997.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r174c02s02	40.0	6189.0	39.0	14.0	688.0	5.0	6975.0
r173c10s01	15.0	5767.0	26.0	19.0	1131.0	0.0	6958.0
r130c12s04	6.0	5831.0	20.0	3.0	1135.0	0.0	6995.0
r162c05s02	69.0	5826.0	35.0	16.0	984.0	1.0	6931.0
r129c12s02	10.0	6190.0	34.0	5.0	709.0	2.0	6950.0
r183c05s01	6.0	1411.0	0.0	10.0	107.0	0.0	1518.0
r129c12s03	12.0	6058.0	47.0	5.0	840.0	2.0	6964.0
r141c09s01	7.0	5945.0	41.0	16.0	977.0	5.0	6991.0
r135c09s03	22.0	5702.0	44.0	16.0	1135.0	1.0	6899.0
r162c05s03	70.0	5650.0	31.0	14.0	1158.0	3.0	6926.0
r136c04s04	18.0	5611.0	31.0	13.0	1280.0	3.0	6936.0
r183c05s02	6.0	2480.0	11.0	12.0	123.0	0.0	2632.0
r174c02s01	3.0	2174.0	13.0	3.0	422.0	0.0	2615.0
r172c02s01	3.0	1259.0	0.0	11.0	270.0	0.0	1529.0
r136c03s01	8.0	5866.0	30.0	10.0	1008.0	0.0	6922.0
r130c12s03	12.0	6601.0	27.0	6.0	329.0	3.0	6978.0
r183c05s03	6.0	2487.0	6.0	16.0	110.0	0.0	2625.0
mean	21.2	4973.4	23.9	11.0	671.1	1.3	5698.2

Tabella 3.78: Risultati Gennaio v2 - Window Mean&sum con errore quadrato medio e alfa 0.3

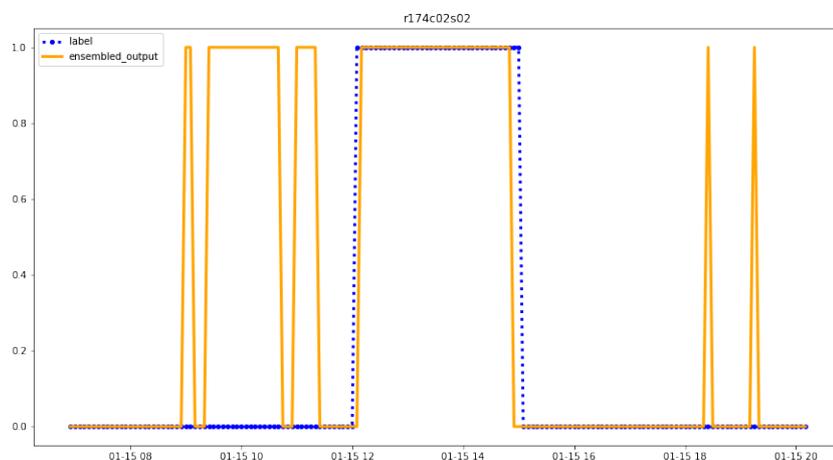


Figura 3.136: Output ensembled model anomalia 15 Gennaio, alfa 0.3, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6889.0	4.0	3.0	19.0	3.0	6924.0
r140c12s04	81.0	6764.0	3.0	7.0	142.0	0.0	6997.0
r174c02s02	40.0	6595.0	23.0	11.0	301.0	5.0	6975.0
r173c10s01	15.0	6339.0	25.0	18.0	561.0	0.0	6958.0
r130c12s04	6.0	6400.0	15.0	3.0	571.0	0.0	6995.0
r162c05s02	67.0	6299.0	29.0	5.0	528.0	3.0	6931.0
r129c12s02	9.0	6531.0	33.0	0.0	374.0	3.0	6950.0
r183c05s01	6.0	1468.0	0.0	2.0	50.0	0.0	1518.0
r129c12s03	10.0	6501.0	44.0	0.0	405.0	4.0	6964.0
r141c09s01	6.0	6464.0	32.0	15.0	468.0	6.0	6991.0
r135c09s03	22.0	6289.0	40.0	16.0	552.0	1.0	6899.0
r162c05s03	67.0	6152.0	24.0	5.0	672.0	6.0	6926.0
r136c04s04	18.0	6223.0	31.0	14.0	667.0	3.0	6936.0
r183c05s02	6.0	2579.0	4.0	2.0	41.0	0.0	2632.0
r174c02s01	2.0	2390.0	13.0	2.0	207.0	1.0	2615.0
r172c02s01	3.0	1411.0	0.0	5.0	118.0	0.0	1529.0
r136c03s01	8.0	6356.0	26.0	3.0	529.0	0.0	6922.0
r130c12s03	12.0	6813.0	5.0	1.0	144.0	3.0	6978.0
r183c05s03	6.0	2576.0	2.0	3.0	38.0	0.0	2625.0
mean	20.5	5317.8	18.6	6.1	336.2	2.0	5698.2

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
--	----	----	--------	---------	-----------	----	---------

Tabella 3.79: Risultati Gennaio v2 - Window Mean&sum con errore quadrato medio e alfa 0.5

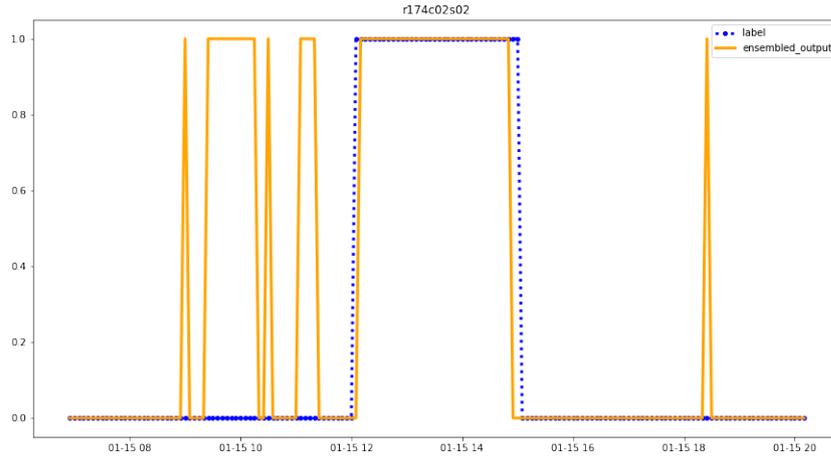


Figura 3.137: Output ensembled model anomalia 15 Gennaio, alfa 0.5, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6910.0	1.0	3.0	1.0	3.0	6924.0
r140c12s04	81.0	6849.0	0.0	7.0	60.0	0.0	6997.0
r174c02s02	40.0	6784.0	16.0	7.0	123.0	5.0	6975.0
r173c10s01	15.0	6645.0	22.0	18.0	258.0	0.0	6958.0
r130c12s04	6.0	6713.0	11.0	3.0	262.0	0.0	6995.0
r162c05s02	65.0	6584.0	14.0	4.0	259.0	5.0	6931.0
r129c12s02	7.0	6771.0	23.0	0.0	144.0	5.0	6950.0
r183c05s01	6.0	1495.0	0.0	2.0	23.0	0.0	1518.0
r129c12s03	9.0	6716.0	33.0	0.0	201.0	5.0	6964.0
r141c09s01	6.0	6740.0	24.0	14.0	201.0	6.0	6991.0
r135c09s03	22.0	6583.0	32.0	16.0	266.0	1.0	6899.0
r162c05s03	65.0	6508.0	10.0	3.0	332.0	8.0	6926.0
r136c04s04	18.0	6616.0	27.0	14.0	278.0	3.0	6936.0
r183c05s02	6.0	2599.0	3.0	1.0	23.0	0.0	2632.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r174c02s01	1.0	2520.0	11.0	2.0	79.0	2.0	2615.0
r172c02s01	3.0	1484.0	0.0	5.0	45.0	0.0	1529.0
r136c03s01	6.0	6653.0	23.0	3.0	235.0	2.0	6922.0
r130c12s03	12.0	6899.0	0.0	0.0	64.0	3.0	6978.0
r183c05s03	6.0	2595.0	2.0	1.0	21.0	0.0	2625.0
mean	20.0	5508.6	13.3	5.4	151.3	2.5	5698.2

Tabella 3.80: Risultati Gennaio v2 - Window Mean&sum con errore quadrato medio e alfa 0.7

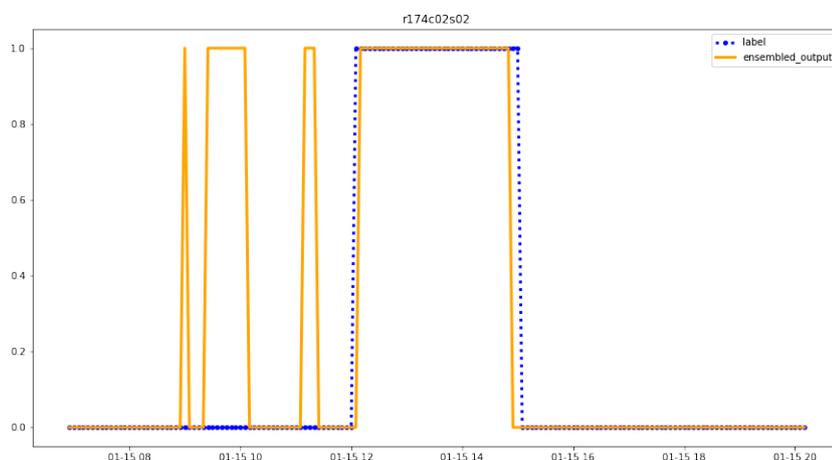


Figura 3.138: Output ensembled model anomalia 15 Gennaio, alfa 0.7, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6896.0	0.0	7.0	13.0	0.0	6997.0
r174c02s02	40.0	6898.0	5.0	6.0	21.0	5.0	6975.0
r173c10s01	15.0	6844.0	13.0	18.0	68.0	0.0	6958.0
r130c12s04	6.0	6901.0	8.0	3.0	77.0	0.0	6995.0
r162c05s02	64.0	6778.0	4.0	2.0	77.0	6.0	6931.0
r129c12s02	7.0	6911.0	4.0	0.0	23.0	5.0	6950.0
r183c05s01	6.0	1505.0	0.0	2.0	13.0	0.0	1518.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r129c12s03	9.0	6913.0	4.0	0.0	33.0	5.0	6964.0
r141c09s01	6.0	6923.0	3.0	7.0	46.0	6.0	6991.0
r135c09s03	22.0	6804.0	24.0	16.0	53.0	1.0	6899.0
r162c05s03	64.0	6746.0	3.0	2.0	102.0	9.0	6926.0
r136c04s04	18.0	6845.0	19.0	13.0	58.0	3.0	6936.0
r183c05s02	6.0	2614.0	1.0	1.0	10.0	0.0	2632.0
r174c02s01	1.0	2594.0	5.0	2.0	11.0	2.0	2615.0
r172c02s01	3.0	1518.0	0.0	5.0	11.0	0.0	1529.0
r136c03s01	5.0	6852.0	11.0	3.0	48.0	3.0	6922.0
r130c12s03	10.0	6941.0	0.0	0.0	22.0	5.0	6978.0
r183c05s03	6.0	2609.0	1.0	1.0	8.0	0.0	2625.0
mean	19.7	5631.8	5.5	4.8	36.5	2.8	5698.2

Tabella 3.81: Risultati Gennaio v2 - Window Mean&sum con errore quadrato medio e alfa 0.9

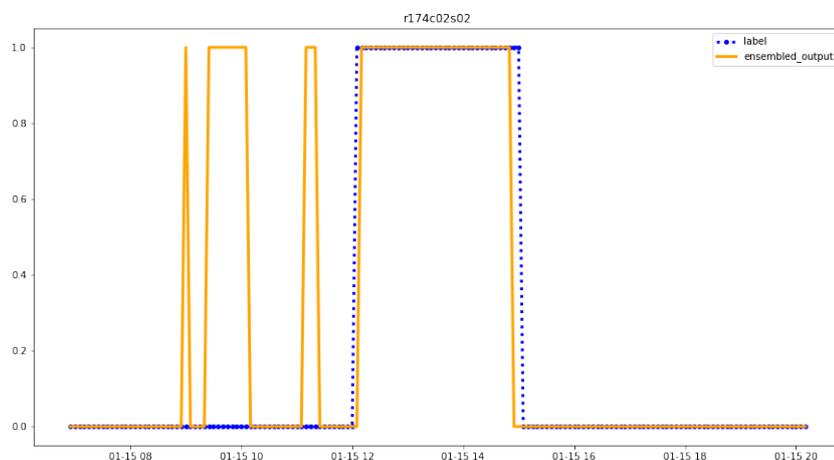


Figura 3.139: Output ensembled model anomalia 15 Gennaio, alfa 0.9, v2-window Mean&sum

Alfa	Anticipo medio
0.3	2 h 19 min
0.5	2 h 05 min
0.7	1 h 40 min
0.9	0 h 53 min

Tabella 3.82: Tabella anticipo medio Gennaio v2 Window Mean & Sum - MSE

Analisi Gennaio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	9.0	6680.0	26.0	5.0	204.0	0.0	6924.0
r140c12s04	81.0	6590.0	33.0	9.0	284.0	0.0	6997.0
r174c02s02	40.0	6618.0	20.0	12.0	280.0	5.0	6975.0
r173c10s01	15.0	6491.0	26.0	19.0	407.0	0.0	6958.0
r130c12s04	6.0	6577.0	14.0	3.0	395.0	0.0	6995.0
r162c05s02	64.0	6432.0	19.0	7.0	403.0	6.0	6931.0
r129c12s02	10.0	6632.0	36.0	4.0	266.0	2.0	6950.0
r183c05s01	6.0	1464.0	0.0	8.0	54.0	0.0	1518.0
r129c12s03	12.0	6587.0	50.0	1.0	312.0	2.0	6964.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r141c09s01	6.0	6502.0	39.0	15.0	423.0	6.0	6991.0
r135c09s03	22.0	6381.0	44.0	16.0	456.0	1.0	6899.0
r162c05s03	64.0	6366.0	14.0	7.0	466.0	9.0	6926.0
r136c04s04	18.0	6474.0	30.0	13.0	418.0	3.0	6936.0
r183c05s02	6.0	2552.0	8.0	2.0	64.0	0.0	2632.0
r174c02s01	3.0	2453.0	12.0	3.0	144.0	0.0	2615.0
r172c02s01	3.0	1446.0	0.0	11.0	83.0	0.0	1529.0
r136c03s01	8.0	6474.0	25.0	10.0	405.0	0.0	6922.0
r130c12s03	12.0	6647.0	33.0	9.0	274.0	3.0	6978.0
r183c05s03	6.0	2544.0	6.0	4.0	65.0	0.0	2625.0
mean	20.6	5363.7	22.9	8.3	284.4	1.9	5698.2

Tabella 3.83: Risultati Gennaio v2 - Window Mean&sum con errore assoluto medio e alfa 0.3

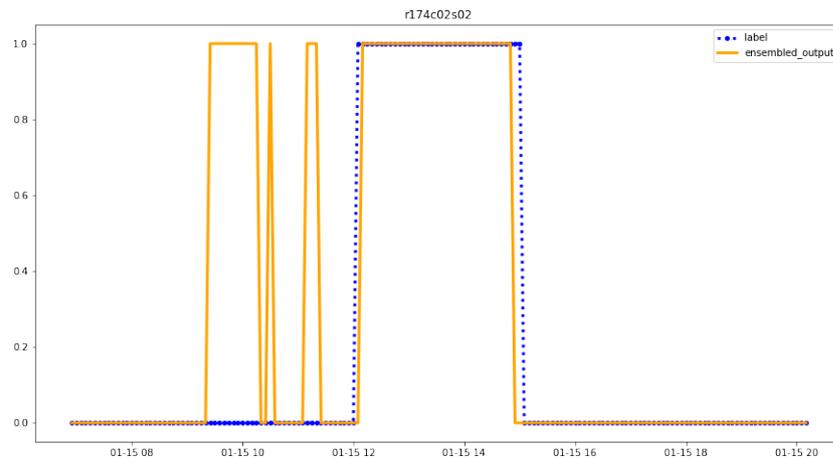


Figura 3.140: Output ensembled model anomalia 15 Gennaio, alfa 0.3, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6865.0	9.0	3.0	38.0	3.0	6924.0
r140c12s04	81.0	6824.0	4.0	7.0	81.0	0.0	6997.0
r174c02s02	40.0	6837.0	15.0	6.0	72.0	5.0	6975.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r173c10s01	15.0	6759.0	23.0	18.0	143.0	0.0	6958.0
r130c12s04	6.0	6812.0	11.0	3.0	163.0	0.0	6995.0
r162c05s02	64.0	6673.0	8.0	5.0	175.0	6.0	6931.0
r129c12s02	7.0	6847.0	20.0	0.0	71.0	5.0	6950.0
r183c05s01	6.0	1498.0	0.0	2.0	20.0	0.0	1518.0
r129c12s03	9.0	6801.0	37.0	0.0	112.0	5.0	6964.0
r141c09s01	6.0	6798.0	25.0	12.0	144.0	6.0	6991.0
r135c09s03	22.0	6686.0	34.0	16.0	161.0	1.0	6899.0
r162c05s03	64.0	6622.0	5.0	5.0	221.0	9.0	6926.0
r136c04s04	18.0	6745.0	29.0	13.0	148.0	3.0	6936.0
r183c05s02	6.0	2604.0	2.0	1.0	19.0	0.0	2632.0
r174c02s01	1.0	2562.0	9.0	2.0	39.0	2.0	2615.0
r172c02s01	3.0	1507.0	0.0	5.0	22.0	0.0	1529.0
r136c03s01	7.0	6753.0	20.0	3.0	138.0	1.0	6922.0
r130c12s03	12.0	6873.0	13.0	1.0	76.0	3.0	6978.0
r183c05s03	6.0	2602.0	2.0	1.0	14.0	0.0	2625.0
mean	19.9	5561.5	14.0	5.4	97.7	2.6	5698.2

Tabella 3.84: Risultati Gennaio v2 - Window Mean&sum con errore assoluto medio e alfa 0.5

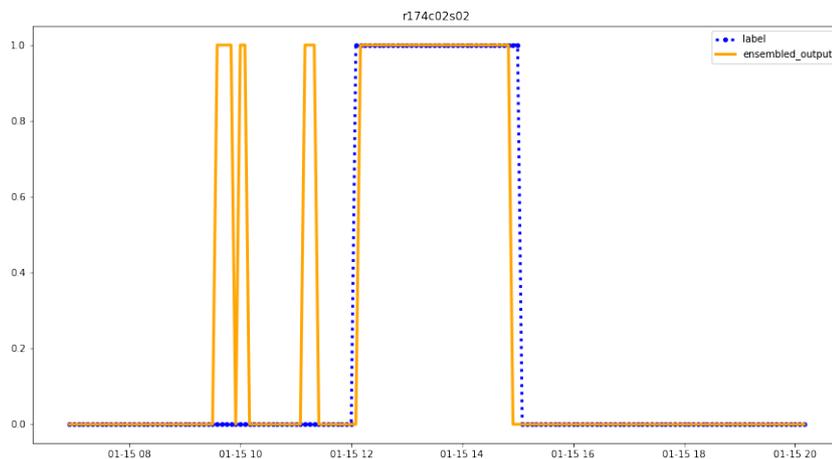


Figura 3.141: Output ensembled model anomalia 15 Gennaio, alfa 0.5, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6910.0	1.0	3.0	1.0	3.0	6924.0
r140c12s04	81.0	6894.0	1.0	7.0	14.0	0.0	6997.0
r174c02s02	40.0	6901.0	4.0	6.0	19.0	5.0	6975.0
r173c10s01	15.0	6868.0	10.0	18.0	47.0	0.0	6958.0
r130c12s04	6.0	6930.0	6.0	3.0	50.0	0.0	6995.0
r162c05s02	64.0	6778.0	3.0	4.0	76.0	6.0	6931.0
r129c12s02	7.0	6923.0	4.0	0.0	11.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6929.0	1.0	0.0	20.0	5.0	6964.0
r141c09s01	6.0	6946.0	0.0	6.0	27.0	6.0	6991.0
r135c09s03	22.0	6823.0	24.0	16.0	34.0	1.0	6899.0
r162c05s03	64.0	6768.0	2.0	4.0	79.0	9.0	6926.0
r136c04s04	18.0	6879.0	19.0	13.0	24.0	3.0	6936.0
r183c05s02	6.0	2624.0	1.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2599.0	1.0	2.0	10.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6877.0	10.0	3.0	24.0	3.0	6922.0
r130c12s03	11.0	6947.0	0.0	0.0	16.0	4.0	6978.0
r183c05s03	6.0	2615.0	2.0	1.0	1.0	0.0	2625.0
mean	19.8	5644.3	4.7	4.9	24.7	2.7	5698.2

Tabella 3.85: Risultati Gennaio v2 - Window Mean&sum con errore assoluto medio e alfa 0.7

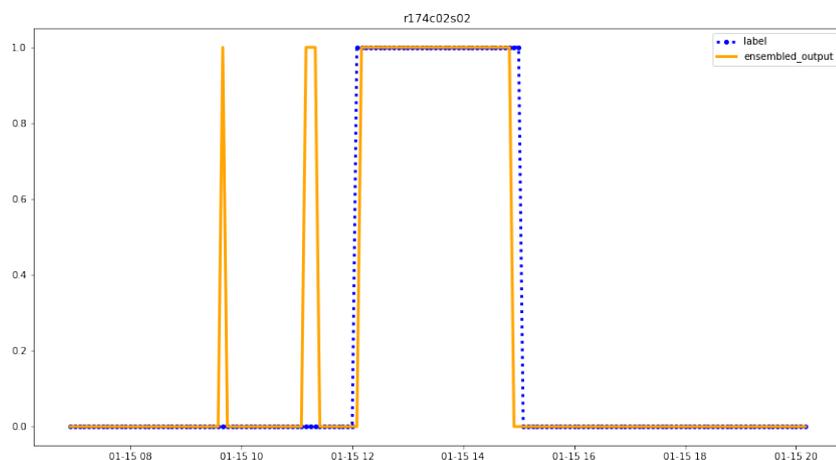


Figura 3.142: Output ensembled model anomalia 15 Gennaio, alfa 0.7, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	6.0	6912.0	0.0	3.0	0.0	3.0	6924.0
r140c12s04	81.0	6909.0	0.0	7.0	0.0	0.0	6997.0
r174c02s02	40.0	6919.0	4.0	6.0	1.0	5.0	6975.0
r173c10s01	15.0	6907.0	1.0	18.0	17.0	0.0	6958.0
r130c12s04	6.0	6985.0	0.0	3.0	1.0	0.0	6995.0
r162c05s02	64.0	6850.0	2.0	2.0	7.0	6.0	6931.0
r129c12s02	7.0	6933.0	1.0	0.0	4.0	5.0	6950.0
r183c05s01	6.0	1510.0	0.0	2.0	8.0	0.0	1518.0
r129c12s03	9.0	6950.0	0.0	0.0	0.0	5.0	6964.0
r141c09s01	6.0	6973.0	0.0	6.0	0.0	6.0	6991.0
r135c09s03	22.0	6870.0	10.0	16.0	1.0	1.0	6899.0
r162c05s03	64.0	6840.0	1.0	2.0	10.0	9.0	6926.0
r136c04s04	18.0	6910.0	11.0	13.0	1.0	3.0	6936.0
r183c05s02	6.0	2625.0	0.0	1.0	0.0	0.0	2632.0
r174c02s01	1.0	2610.0	0.0	2.0	0.0	2.0	2615.0
r172c02s01	3.0	1521.0	0.0	5.0	8.0	0.0	1529.0
r136c03s01	5.0	6910.0	1.0	3.0	0.0	3.0	6922.0
r130c12s03	10.0	6962.0	0.0	0.0	1.0	5.0	6978.0
r183c05s03	6.0	2618.0	0.0	1.0	0.0	0.0	2625.0
mean	19.7	5669.2	1.6	4.7	3.1	2.8	5698.2

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
--	----	----	--------	---------	-----------	----	---------

Tabella 3.86: Risultati Gennaio v2 - Window Mean&sum con errore assoluto medio e alfa 0.9

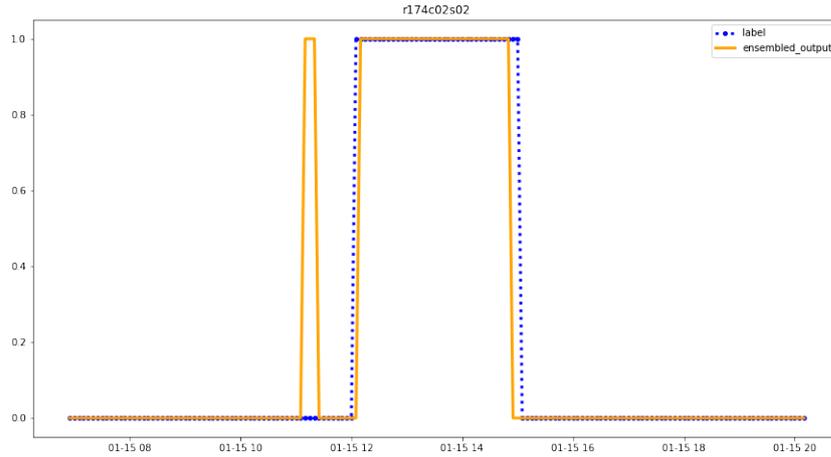


Figura 3.143: Output ensembled model anomalia 15 Gennaio, alfa 0.9, v2-window Mean&sum

Alfa	Anticipo medio
0.3	2 h 15 min
0.5	1 h 42 min
0.7	0 h 37 min
0.9	0 h 33 min

Tabella 3.87: Tabella anticipo medio Gennaio v2 Window Mean & Sum - MAE

Analisi Maggio errore quadrato medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	5581.0	13.0	5.0	1514.0	3.0	7118.0
r141c09s02	1.0	5869.0	13.0	0.0	1326.0	1.0	7210.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r140c12s04	4.0	6893.0	16.0	5.0	290.0	1.0	7209.0
r174c02s02	5.0	5657.0	8.0	4.0	1538.0	0.0	7212.0
r173c10s01	5.0	5855.0	17.0	12.0	1266.0	0.0	7155.0
r130c12s04	110.0	5823.0	29.0	4.0	1231.0	1.0	7198.0
r162c05s02	1281.0	4476.0	137.0	85.0	1119.0	91.0	7165.0
r129c12s02	2.0	5654.0	24.0	0.0	1443.0	0.0	7123.0
r183c05s01	4.0	6010.0	10.0	12.0	1169.0	0.0	7205.0
r129c12s03	2.0	5824.0	15.0	2.0	1238.0	0.0	7081.0
r141c09s01	1.0	5772.0	21.0	0.0	1412.0	1.0	7207.0
r135c09s03	0.0	5818.0	20.0	0.0	1304.0	2.0	7144.0
r162c05s03	1350.0	4326.0	129.0	104.0	1307.0	26.0	7165.0
r136c04s04	2.0	5470.0	10.0	5.0	1668.0	0.0	7155.0
r183c05s02	862.0	5537.0	18.0	1.0	777.0	3.0	7198.0
r174c02s01	6.0	6071.0	15.0	1.0	1118.0	2.0	7213.0
r172c02s01	5.0	5621.0	17.0	5.0	1540.0	0.0	7188.0
r136c03s01	1.0	5768.0	13.0	2.0	1338.0	1.0	7123.0
r130c12s03	111.0	5915.0	28.0	24.0	1122.0	1.0	7201.0
r183c05s03	7.0	5791.0	15.0	9.0	1387.0	0.0	7209.0
mean	188.1	5686.6	28.4	14.0	1255.3	6.7	7173.9

Tabella 3.88: Risultati Maggio v2 - Window Mean&sum con errore quadrato medio e alfa 0.3

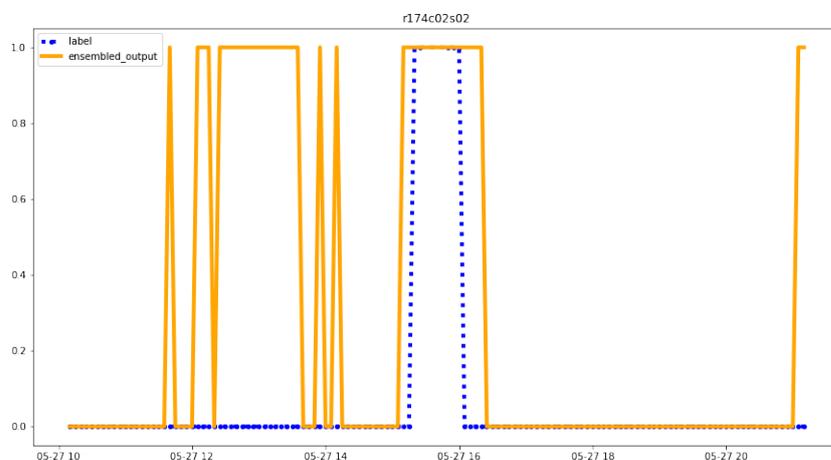


Figura 3.144: Output ensembled model anomalia 27 Maggio, alfa 0.3, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	6248.0	9.0	4.0	852.0	3.0	7118.0
r141c09s02	1.0	6542.0	8.0	0.0	658.0	1.0	7210.0
r140c12s04	2.0	7097.0	7.0	1.0	99.0	3.0	7209.0
r174c02s02	5.0	6364.0	4.0	1.0	838.0	0.0	7212.0
r173c10s01	5.0	6498.0	9.0	9.0	634.0	0.0	7155.0
r130c12s04	110.0	6392.0	19.0	3.0	673.0	1.0	7198.0
r162c05s02	1279.0	5005.0	119.0	74.0	619.0	93.0	7165.0
r129c12s02	2.0	6333.0	18.0	0.0	770.0	0.0	7123.0
r183c05s01	4.0	6625.0	10.0	2.0	564.0	0.0	7205.0
r129c12s03	2.0	6435.0	14.0	0.0	630.0	0.0	7081.0
r141c09s01	1.0	6448.0	16.0	0.0	741.0	1.0	7207.0
r135c09s03	0.0	6487.0	15.0	0.0	640.0	2.0	7144.0
r162c05s03	1344.0	4819.0	116.0	100.0	831.0	32.0	7165.0
r136c04s04	1.0	6260.0	7.0	3.0	883.0	1.0	7155.0
r183c05s02	857.0	5986.0	2.0	1.0	344.0	8.0	7198.0
r174c02s01	6.0	6599.0	7.0	1.0	598.0	2.0	7213.0
r172c02s01	2.0	6356.0	12.0	0.0	815.0	3.0	7188.0
r136c03s01	1.0	6413.0	10.0	2.0	696.0	1.0	7123.0
r130c12s03	111.0	6439.0	23.0	24.0	603.0	1.0	7201.0
r183c05s03	5.0	6505.0	9.0	1.0	687.0	2.0	7209.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
mean	187.0	6292.6	21.7	11.3	658.8	7.7	7173.9

Tabella 3.89: Risultati Maggio v2 - Window Mean&sum con errore quadrato medio e alfa 0.5

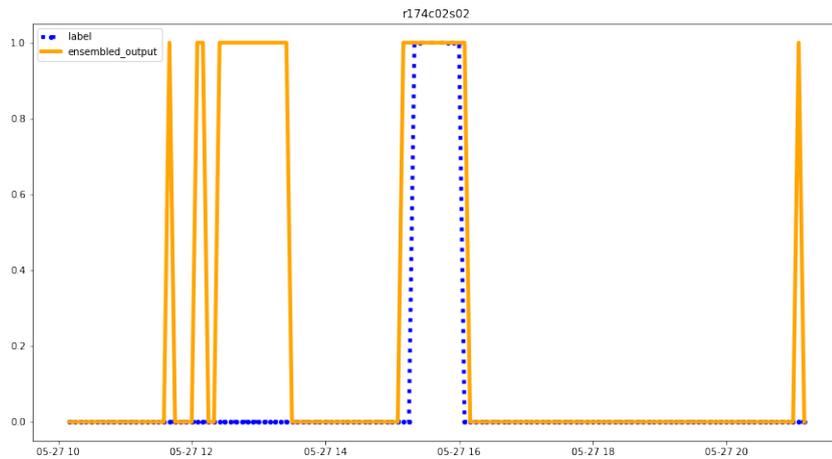


Figura 3.145: Output ensembled model anomalia 27 Maggio, alfa 0.5, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	6701.0	7.0	1.0	404.0	3.0	7118.0
r141c09s02	1.0	6890.0	5.0	0.0	313.0	1.0	7210.0
r140c12s04	2.0	7152.0	3.0	1.0	48.0	3.0	7209.0
r174c02s02	4.0	6785.0	3.0	0.0	419.0	1.0	7212.0
r173c10s01	4.0	6847.0	2.0	5.0	296.0	1.0	7155.0
r130c12s04	110.0	6719.0	10.0	2.0	356.0	1.0	7198.0
r162c05s02	1278.0	5231.0	108.0	69.0	409.0	94.0	7165.0
r129c12s02	2.0	6757.0	12.0	0.0	352.0	0.0	7123.0
r183c05s01	1.0	6937.0	10.0	1.0	253.0	3.0	7205.0
r129c12s03	2.0	6766.0	12.0	0.0	301.0	0.0	7081.0
r141c09s01	1.0	6869.0	9.0	0.0	327.0	1.0	7207.0
r135c09s03	0.0	6804.0	11.0	0.0	327.0	2.0	7144.0
r162c05s03	1344.0	5071.0	105.0	96.0	594.0	32.0	7165.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c04s04	1.0	6760.0	5.0	2.0	386.0	1.0	7155.0
r183c05s02	857.0	6217.0	0.0	1.0	115.0	8.0	7198.0
r174c02s01	6.0	6901.0	5.0	1.0	298.0	2.0	7213.0
r172c02s01	0.0	6766.0	5.0	0.0	412.0	5.0	7188.0
r136c03s01	1.0	6823.0	5.0	1.0	292.0	1.0	7123.0
r130c12s03	111.0	6739.0	13.0	24.0	313.0	1.0	7201.0
r183c05s03	4.0	6924.0	5.0	0.0	273.0	3.0	7209.0
mean	186.6	6632.9	16.8	10.2	324.4	8.2	7173.9

Tabella 3.90: Risultati Maggio v2 - Window Mean&sum con errore quadrato medio e alfa 0.7

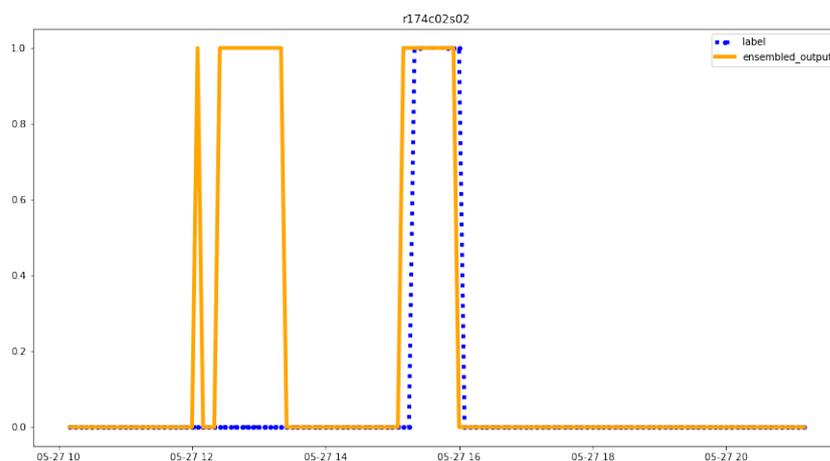


Figura 3.146: Output ensembled model anomalia 27 Maggio, alfa 0.7, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7022.0	0.0	0.0	91.0	3.0	7118.0
r141c09s02	1.0	7128.0	0.0	0.0	80.0	1.0	7210.0
r140c12s04	2.0	7196.0	0.0	1.0	7.0	3.0	7209.0
r174c02s02	4.0	7060.0	2.0	0.0	145.0	1.0	7212.0
r173c10s01	2.0	7077.0	1.0	4.0	68.0	3.0	7155.0
r130c12s04	110.0	6970.0	5.0	1.0	111.0	1.0	7198.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r162c05s02	1278.0	5399.0	93.0	68.0	257.0	94.0	7165.0
r129c12s02	2.0	7046.0	4.0	0.0	71.0	0.0	7123.0
r183c05s01	1.0	7151.0	1.0	0.0	49.0	3.0	7205.0
r129c12s03	2.0	6999.0	6.0	0.0	74.0	0.0	7081.0
r141c09s01	1.0	7118.0	1.0	0.0	86.0	1.0	7207.0
r135c09s03	0.0	7066.0	1.0	0.0	75.0	2.0	7144.0
r162c05s03	1343.0	5235.0	97.0	93.0	441.0	33.0	7165.0
r136c04s04	1.0	7042.0	3.0	1.0	107.0	1.0	7155.0
r183c05s02	857.0	6315.0	0.0	1.0	17.0	8.0	7198.0
r174c02s01	6.0	7094.0	4.0	1.0	106.0	2.0	7213.0
r172c02s01	0.0	7070.0	1.0	0.0	112.0	5.0	7188.0
r136c03s01	1.0	7057.0	1.0	0.0	63.0	1.0	7123.0
r130c12s03	111.0	6985.0	1.0	24.0	79.0	1.0	7201.0
r183c05s03	4.0	7151.0	0.0	0.0	51.0	3.0	7209.0
mean	186.4	6859.1	11.1	9.7	104.5	8.3	7173.9

Tabella 3.91: Risultati Maggio v2 - Window Mean&sum con errore quadrato medio e alfa 0.9

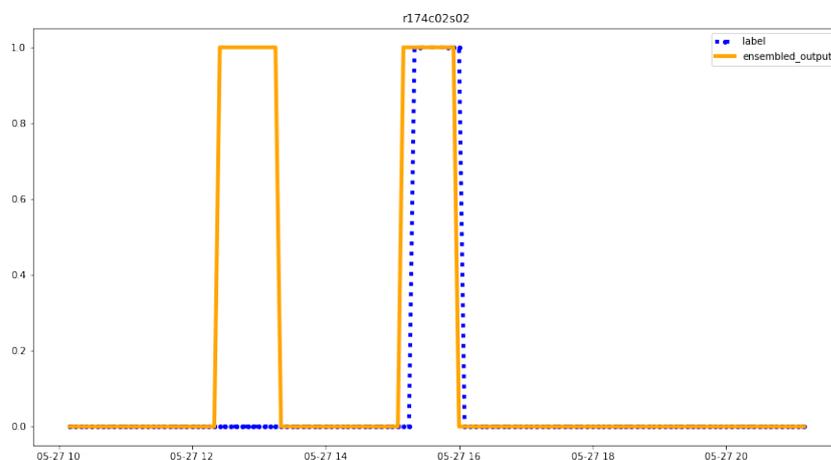


Figura 3.147: Output ensembled model anomalia 27 Maggio, alfa 0.9, v2-window Mean&sum

Alfa	Anticipo medio
0.3	1 h 35 min
0.5	1 h 25 min
0.7	1 h 12 min
0.9	1 h 00 min

Tabella 3.92: Tabella anticipo medio Maggio v2 Window Mean & Sum - MSE

Analisi Maggio errore assoluto medio

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	6482.0	11.0	3.0	617.0	3.0	7118.0
r141c09s02	1.0	6735.0	4.0	0.0	469.0	1.0	7210.0
r140c12s04	2.0	6950.0	20.0	1.0	233.0	3.0	7209.0
r174c02s02	5.0	6655.0	6.0	2.0	544.0	0.0	7212.0
r173c10s01	4.0	6694.0	6.0	10.0	440.0	1.0	7155.0
r130c12s04	110.0	6520.0	28.0	1.0	538.0	1.0	7198.0
r162c05s02	1283.0	5044.0	125.0	78.0	570.0	89.0	7165.0
r129c12s02	2.0	6628.0	23.0	0.0	470.0	0.0	7123.0
r183c05s01	3.0	6788.0	10.0	2.0	401.0	1.0	7205.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r129c12s03	2.0	6593.0	13.0	3.0	470.0	0.0	7081.0
r141c09s01	1.0	6668.0	15.0	0.0	522.0	1.0	7207.0
r135c09s03	0.0	6635.0	19.0	0.0	488.0	2.0	7144.0
r162c05s03	1343.0	4896.0	119.0	95.0	756.0	33.0	7165.0
r136c04s04	1.0	6601.0	6.0	1.0	545.0	1.0	7155.0
r183c05s02	862.0	5942.0	8.0	1.0	382.0	3.0	7198.0
r174c02s01	6.0	6747.0	15.0	1.0	442.0	2.0	7213.0
r172c02s01	2.0	6650.0	12.0	0.0	521.0	3.0	7188.0
r136c03s01	1.0	6663.0	11.0	0.0	447.0	1.0	7123.0
r130c12s03	111.0	6531.0	29.0	24.0	505.0	1.0	7201.0
r183c05s03	5.0	6711.0	11.0	6.0	474.0	2.0	7209.0
mean	187.3	6456.6	24.6	11.4	491.7	7.4	7173.9

Tabella 3.93: Risultati Maggio v2 - Window Mean&sum con errore assoluto medio e alfa 0.3

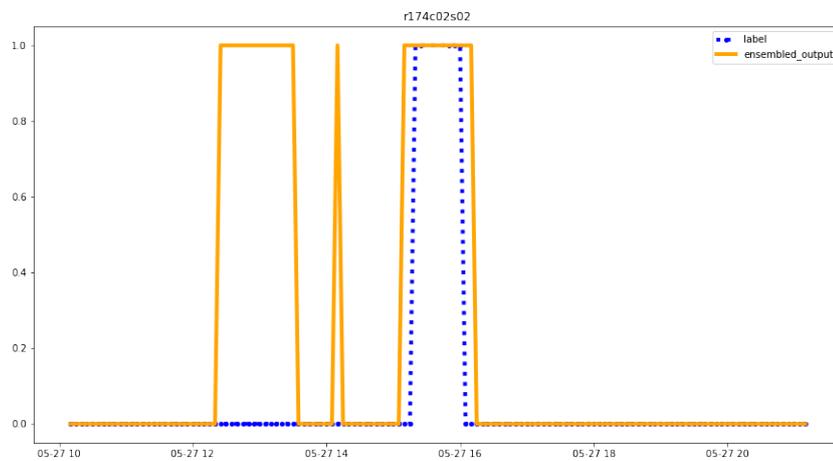


Figura 3.148: Output ensembled model anomalia 27 Maggio, alfa 0.3, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	6900.0	7.0	1.0	205.0	3.0	7118.0
r141c09s02	1.0	7051.0	0.0	0.0	157.0	1.0	7210.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r140c12s04	2.0	7149.0	6.0	1.0	48.0	3.0	7209.0
r174c02s02	4.0	6984.0	3.0	0.0	220.0	1.0	7212.0
r173c10s01	3.0	6968.0	1.0	5.0	176.0	2.0	7155.0
r130c12s04	110.0	6854.0	12.0	1.0	220.0	1.0	7198.0
r162c05s02	1278.0	5321.0	107.0	69.0	320.0	94.0	7165.0
r129c12s02	2.0	6980.0	12.0	0.0	129.0	0.0	7123.0
r183c05s01	1.0	7062.0	3.0	0.0	136.0	3.0	7205.0
r129c12s03	2.0	6911.0	13.0	0.0	155.0	0.0	7081.0
r141c09s01	1.0	7001.0	7.0	0.0	197.0	1.0	7207.0
r135c09s03	0.0	6936.0	9.0	0.0	197.0	2.0	7144.0
r162c05s03	1343.0	5140.0	108.0	93.0	525.0	33.0	7165.0
r136c04s04	1.0	6964.0	4.0	1.0	184.0	1.0	7155.0
r183c05s02	857.0	6240.0	0.0	1.0	92.0	8.0	7198.0
r174c02s01	6.0	6997.0	6.0	1.0	201.0	2.0	7213.0
r172c02s01	0.0	6980.0	3.0	0.0	200.0	5.0	7188.0
r136c03s01	1.0	6969.0	5.0	0.0	147.0	1.0	7123.0
r130c12s03	111.0	6860.0	9.0	24.0	196.0	1.0	7201.0
r183c05s03	4.0	7048.0	3.0	0.0	151.0	3.0	7209.0
mean	186.4	6765.8	15.9	9.8	192.8	8.2	7173.9

Tabella 3.94: Risultati Maggio v2 - Window Mean&sum con errore assoluto medio e alfa 0.5

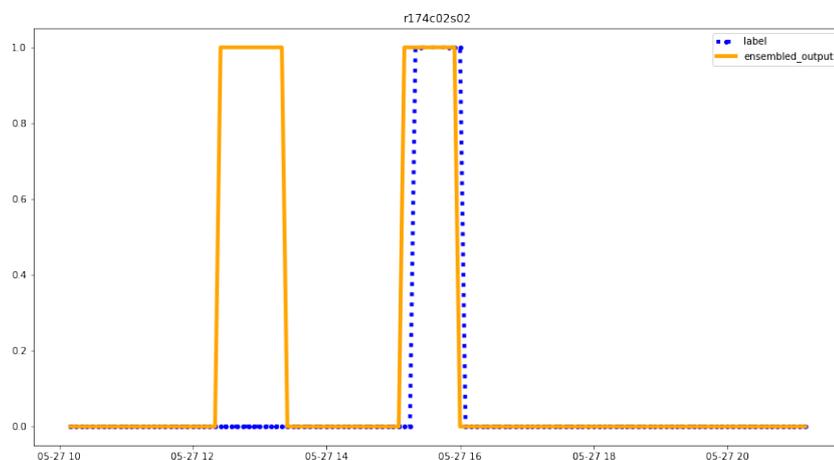


Figura 3.149: Output ensemble model anomalia 27 Maggio, alfa 0.5, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7072.0	0.0	0.0	41.0	3.0	7118.0
r141c09s02	1.0	7188.0	0.0	0.0	20.0	1.0	7210.0
r140c12s04	2.0	7191.0	1.0	1.0	11.0	3.0	7209.0
r174c02s02	4.0	7136.0	2.0	0.0	69.0	1.0	7212.0
r173c10s01	2.0	7115.0	0.0	4.0	31.0	3.0	7155.0
r130c12s04	110.0	7015.0	5.0	1.0	66.0	1.0	7198.0
r162c05s02	1278.0	5406.0	95.0	68.0	248.0	94.0	7165.0
r129c12s02	2.0	7098.0	5.0	0.0	18.0	0.0	7123.0
r183c05s01	1.0	7178.0	1.0	0.0	22.0	3.0	7205.0
r129c12s03	2.0	7043.0	7.0	0.0	29.0	0.0	7081.0
r141c09s01	1.0	7162.0	1.0	0.0	42.0	1.0	7207.0
r135c09s03	0.0	7107.0	1.0	0.0	34.0	2.0	7144.0
r162c05s03	1343.0	5254.0	98.0	92.0	422.0	33.0	7165.0
r136c04s04	1.0	7101.0	3.0	0.0	49.0	1.0	7155.0
r183c05s02	857.0	6325.0	0.0	1.0	7.0	8.0	7198.0
r174c02s01	6.0	7139.0	4.0	1.0	61.0	2.0	7213.0
r172c02s01	0.0	7119.0	1.0	0.0	63.0	5.0	7188.0
r136c03s01	1.0	7099.0	1.0	0.0	21.0	1.0	7123.0
r130c12s03	111.0	7016.0	1.0	24.0	48.0	1.0	7201.0
r183c05s03	4.0	7174.0	0.0	0.0	28.0	3.0	7209.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
mean	186.4	6896.9	11.3	9.6	66.5	8.3	7173.9

Tabella 3.95: Risultati Maggio v2 - Window Mean&sum con errore assoluto medio e alfa 0.7

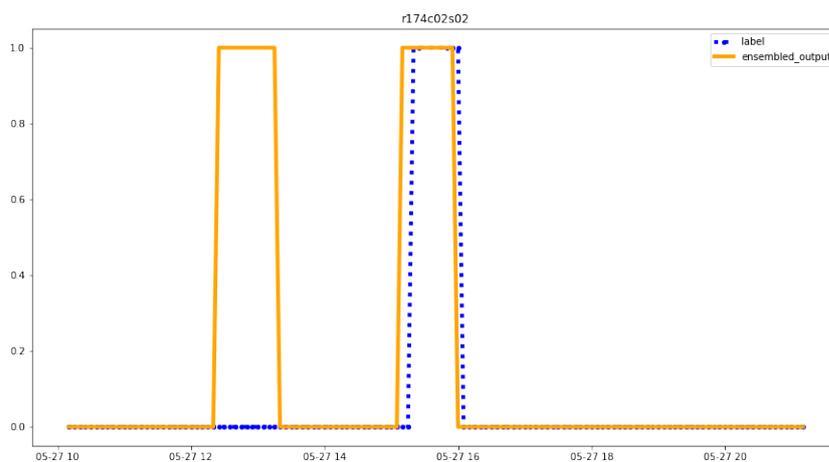


Figura 3.150: Output ensembled model anomalia 27 Maggio, alfa 0.7, v2-window Mean&sum

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c03s02	2.0	7112.0	0.0	0.0	1.0	3.0	7118.0
r141c09s02	1.0	7208.0	0.0	0.0	0.0	1.0	7210.0
r140c12s04	2.0	7203.0	0.0	1.0	0.0	3.0	7209.0
r174c02s02	4.0	7204.0	2.0	0.0	1.0	1.0	7212.0
r173c10s01	2.0	7146.0	0.0	4.0	0.0	3.0	7155.0
r130c12s04	110.0	7081.0	5.0	1.0	0.0	1.0	7198.0
r162c05s02	1278.0	5425.0	91.0	68.0	233.0	94.0	7165.0
r129c12s02	2.0	7120.0	1.0	0.0	0.0	0.0	7123.0
r183c05s01	1.0	7200.0	1.0	0.0	0.0	3.0	7205.0
r129c12s03	2.0	7079.0	0.0	0.0	0.0	0.0	7081.0
r141c09s01	1.0	7203.0	0.0	0.0	2.0	1.0	7207.0
r135c09s03	0.0	7142.0	0.0	0.0	0.0	2.0	7144.0
r162c05s03	1343.0	5281.0	96.0	92.0	397.0	33.0	7165.0

Continua nella prossima pagina

	tp	tn	fp_pre	fp_post	fp_random	fn	records
r136c04s04	1.0	7151.0	2.0	0.0	0.0	1.0	7155.0
r183c05s02	857.0	6330.0	0.0	1.0	2.0	8.0	7198.0
r174c02s01	6.0	7202.0	0.0	1.0	2.0	2.0	7213.0
r172c02s01	0.0	7180.0	0.0	0.0	3.0	5.0	7188.0
r136c03s01	1.0	7120.0	1.0	0.0	0.0	1.0	7123.0
r130c12s03	111.0	7061.0	1.0	24.0	3.0	1.0	7201.0
r183c05s03	4.0	7193.0	0.0	0.0	9.0	3.0	7209.0
mean	186.4	6932.1	10.0	9.6	32.6	8.3	7173.9

Tabella 3.96: Risultati Maggio v2 - Window Mean&sum con errore assoluto medio e alfa 0.9

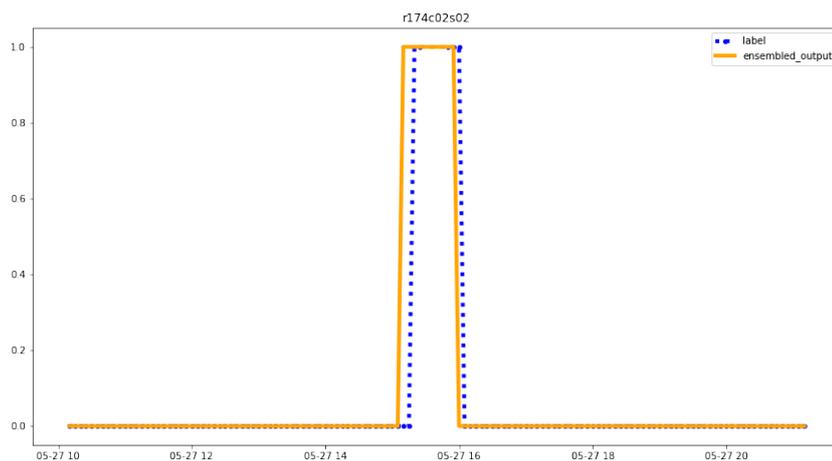


Figura 3.151: Output ensembled model anomalia 27 Maggio, alfa 0.9, v2-window Mean&sum

La proprietà più evidente di questo modello è che al crescere di alfa diminuiscono il numero di falsi positivi, questo è riconducibile al fatto che EMA da più peso agli ultimi valori dell'errore e quindi si verifica con meno frequenza la condizione necessaria alla detection. Ciò non è necessariamente un male poiché fornisce un parametro regolabile in base alle esigenze.

La presenza di più falsi positivi pre-anomalia e anche post, rispetto al modello v2 Window Mean, può essere vista anche nei plot alla fine di ogni tabella.

Alfa	Anticipo medio
0.3	1 h 29 min
0.5	1 h 11 min
0.7	1 h 00 min
0.9	0 h 57 min

Tabella 3.97: Tabella anticipo medio Maggio v2 Window Mean & Sum - MAE

L'utilizzo di MSE in generale presenta un tasso di falsi positivi più alto rispetto all'utilizzo di MAE.

Un'altra proprietà è quella che utilizzando MAE e alfa pari a 0.9 si hanno le stesse performance di un modello MAE che utilizza la convoluzione con un filtro di media.

3.4 Confronto

Metodo	tp	tn	fp_pre	fp_post	fp_random	fn
Random forest	20.4	5675.5	1.1	1.2	0.4	2.2
AE + classificatore supervised	19.7	5669.6	1.5	4.7	2.9	2.8
AE semi-supervised (soglia fissa)	7.4	5489.0	4.89	12.52	182.15	15.1
AE semi-supervised (conv + EMA)	11.5	2847.9	36.7	14.9	2779.3	11.0
Modello ensembled v2 - WM- MAE	19.7	5669.5	1.5	4.7	2.9	2.8
Modello ensembled v2 - WM - MSE	19.7	5668.5	1.9	4.7	3.5	2.8
Modello ensembled v2 - WM - SUM	19.8	5644.3	4.7	4.9	24.7	2.7

Tabella 3.98: Tabella di confronto tecniche dataset Gennaio

Metodo	tp	tn	fp_pre	fp_post	fp_random	fn
Random forest	191.6	6978.9	2.0	2.0	0	3.2
AE + classificatore supervised	186.4	6932.3	10.0	9.6	32.5	8.3
AE semi-supervised (soglia fissa)	47.9	6732.4	4.1	3.9	229.6	146.9
AE semi-supervised (conv + EMA)	95.6	3530.6	30.8	11.95	3409.2	99.1
Modello ensembled v2 - WM-MAE	186.4	6932.2	10.0	9.6	32.5	8.3
Modello ensembled v2 - WM-MSE	186.4	6930.1	10.2	9.6	34.4	8.3
Modello ensembled v2 - WM - SUM	186.4	6896.9	11.3	9.6	66.5	8.3

Tabella 3.99: Tabella di confronto tecniche dataset Maggio

In questa sezione confronterò gli ultimi modelli esposti con versioni intermedie degli stessi, che utilizzano tecniche di detection differenti e l'algoritmo Random Forest di scikit-learn.

Al fine di rendere le tabelle (Tabella 3.98 e Tabella 3.99) il più compatte e leggibili possibile, per il modello ensembled v2 e AE semi-supervised con soglia EMA è stato utilizzato alfa pari a 0.7.

Il modello Random Forest è stato utilizzato con i parametri di default, senza effettuare alcuna operazione di ottimizzazione di quest'ultimi poiché sarebbe risultata in istanze di Random Forest con parametri differenti per ogni nodo e avrebbe reso il confronto non del tutto equo nei confronti dei modelli dotati di autoencoder, la quale struttura è fissa per ogni nodo.

I risultati riportati nelle tabelle sono le medie dei risultati ottenuti effettuando gli esperimenti sui vari nodi analizzati nelle sezioni precedenti.

Si può osservare un aspetto molto importante, per quanto riguarda "Modello ensembled v2 - WM - MAE" e "AE + classificatore supervised", hanno esattamente gli stessi valori. Questo significa che in caso di utilizzo dell'errore assoluto AE-semisupervised non interviene mai ed è anche il motivo perché i risultati con MAE sono identici, mentre nel caso di errore quadrato si può vedere come il tasso di falsi positivi pre-anomalia sia leggermente più alto nel caso di alfa pari a 0.7, ma incrementa col decrescere di alfa.

Per le performance del modello che fa uso della somma sono state prese come riferimento quelle che utilizzano MAE e alfa 0.7, come detto in precedenza con questo setup più alfa incrementa più si va verso la performance di AE + classificatore.

Ma nel caso si utilizzi MSE e un alfa minore di 0.9 si ha un tasso di falsi positivi più alto in generale.

L'algoritmo Random Forest è il chiaro vincitore in termini di tp e fp_random nel confronto appena effettuato, se l'attività appena svolta trattasse solamente la detection di un'anomalia sarebbe sicuramente la tecnica da scegliere. Ma in termini di falsi positivi pre-anomalia è abbastanza scarso poiché non li considera come un pro durante la propria fase di costruzione.

Capitolo 4

Conclusioni

In quest'attività si è andato molto in profondità in termini di analisi delle anomalie, cercando di tenere comunque lo spettro di nodi analizzati e dei periodi abbastanza ampio.

In generale si è evidenziato come gli approcci supervised hanno delle limitazioni quando si vogliono effettuare rilevazioni in anticipo di un'anomalia in caso di dataset sbilanciato e come sia necessario introdurre un ulteriore componente non allenato in modo supervised.

Il modello semisupervised non dando come output direttamente una label, ma un errore di ricostruzione, necessita di ulteriori passi e manipolazioni del proprio output per poter ricavare una label.

È stato analizzato facendo una ricerca abbastanza esaustiva come EMA singolarmente non sia una soglia particolarmente affidabile, almeno in questo contesto, in quanto fornisce il segnale di anomalia non appena l'errore ha un andamento crescente e nel nostro caso l'errore ha un comportamento fortemente oscillatorio.

Date le proprietà del modello supervised e di quello semisupervised si è provato a regolarizzare questo alto tasso di falsi positivi, utilizzando come valore di riferimento la media che nella versione dove si utilizza la convoluzione con filtro di media risulta in un'esclusione quasi completa del modello semisupervised.

Nel caso invece si utilizzi la somma dell'errore degli ultimi 24 record si ha un grosso abbattimento dei falsi positivi random che riportava il modello senza regolarizzazione mantenendo i falsi positivi pre-anomalia a dei valori accettabili, includendo comunque un'opzione di tuning del tasso di falsi positivi tramite il parametro alfa di EMA.

In generale i risultati ottenuti, soprattutto con l'ultimo modello non sono male, nonostante si tratti di un sistema in production nel quale le anomalie sono tutt'altro che frequenti.

Sicuramente questo studio è un altro punto di partenza in termini di ricerca in quanto potrebbero essere esplorate soluzioni differenti per quanto riguarda la fase di rilevazione per AE-semisupervised.

Bibliografia

- [1] Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, Nitesh V. Chawla, A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data, 2018.
- [2] Yixing Wang, Meiqin Liu, Zhejing Bao, Senlin Zhang, Stacked sparse autoencoder with PCA and SVM for data-based line trip fault diagnosis in power systems, 2018.
- [3] Elisabeth Baseman, Sean Blanchard, Nathan DeBardeleben, Amanda Bonnie, Adam Morrow, Interpretable Anomaly Detection for Monitoring of High Performance Computing Systems.
- [4] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano and L. Benini, A Semi-supervised Autoencoder-based Approach for Anomaly Detection in High Performance Computing Systems.
- [5] Raghavendra Chalapathy, Deep Learning for Anomaly Detection: A Survey, 2019.
- [6] Dor Bank, Noam Koenigstein, Raja Giryes, Autoencoders, 2020.
- [7] Francois Chollet, Building autoencoders in Keras, <https://blog.keras.io/building-autoencoders-in-keras.html>
- [8] Kevin Leto. Anomaly Detection in HPC Systems. Tesi di laurea, Università degli Studi di Bologna, 2019.
- [9] <https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/>
- [10] Gian Antonio Susto, Jian Wan, Simone Pampuri, Mattia Zanon. An adaptive machine learning decision system for flexible predictive maintenance 2014.

- [11] Lorenzo Pizzigati, Anomaly Prediction with Temporal Convolutional Networks for HPC Systems.