

Sistemi di supporto alle decisioni

M9Museum Project

Svolto da: Achilli Mattia, Rettaroli Andrea.

Obbiettivo

L'obbiettivo di questo progetto è quello di riuscire a fare delle previsioni sulle serie storiche dell'M9Museum di Venezia. Le serie storiche che compongono il dataset monitorano l'affluenza al museo nelle varie sale. Tramite modelli statistici e modelli neurali si vuole prevedere l'afflusso di visitatori al museo.

Analisi

Analisi del problema

Analisi dei dati

Al fine di riuscire ad applicare i modelli statistici e i modelli neurali, risulta di cruciale importanza analizzare e comprendere i dati in nostro possesso. In seguito si riporta la struttura del file CSV.

timestamp,"area_code","totals","alarms","date"
1632669666838,"floor_3","0","0","2021-09-26"
1632669667149,"front_desk","0","0","2021-09-26"
1632669667226,"floor_1","17","0","2021-09-26"
1632669667600,"floor_2","9","0","2021-09-26"

Dal CSV si individuano i dati timestamp, area_code e totals come i dati di rilievo su cui si costruisce la serie storica del numero di presenti nelle varie sale. Il timestamp indica il momento temporale in cui il dato è stato acquisito. il seguente esempio ci aiuta a capire il formato:

Epoch timestamp: 1640692068

Timestamp in milliseconds: 1640692068000

Date and time (GMT): Tuesday 28 December 2021 11:47:48

Date and time (your time zone): martedì 28 dicembre 2021
12:47:48 GMT+01:00

A seguito di questa conversione si nota che le misurazioni avvengono ogni circa 2secondi. Si decide che per costruire la serie si prendono i riferimenti ogni 10 minuti. Il grafico seguente mostra la serie costruita dall'intero dataset.

AGGIUNGERE IMMAGINE GENERALE.PNG

In seguito si vede il grafico utilizzando i riferimenti ogni 10minuti.

AGGIUNGERE IMMAGINE 3florCorrect-ordered-totals.png

Dal secondo grafico è facilmente rilevabile che all'interno della settimana vi sono dei giorni in cui l'afflusso è 0, ci siamo interfacciati con il Sig.re Luca Agatensi, coordinatore del progetto, per capire se questi dati fossero dovuti a delle chiusure causate dal Covid19 o a degli errori di misurazione o ad altro. In quei giorni il Museo non è aperto al pubblico, ciò prova la correttezza dei dati letti. La serie risulta continua, non si rileva la presenza di outlier.

Ci viene esplicitamente richiesto di ignorare il front_desk in quanto relativo al personale.

In questa fase si leggono i dati da file, essendo 11.4M i record si pensa che sia meglio passare tramite DB e ottenerli già filtrati per sala al fine di ottimizzare i tempi.

Scelte implementative

