

Università degli studi di Bari facoltà di
scienze MM.FF.NN

Progetto ingegneria della conoscenza

TrainDelay-project

by

Vito Proscia mat. 735975



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

Anno accademico 2022-2023

Contents

1	Introduzione	3
1.1	Contesto	3
1.2	Definizione obiettivo principale	3
1.3	Tool utilizzati	3
2	Rappresentazione formale della conoscenza	3
3	Analisi del dataset	5
3.1	Descrizione features	5
3.2	Preparazione dati	5
4	Machine Learning	5
5	Valutazione dei modelli	5
5.1	Metriche scelte	5
6	Risultati	5
7	Sviluppi futuri	5

1 Introduzione

1.1 Contesto

TrainDelay-project è

1.2 Definizione obiettivo principale

L'obiettivo principale del progetto è la creazione di un motore di ricerca che trova i migliori itinerari di viaggio in treno sulla base della stazione di partenza e di arrivo e che, per ogni viaggio, mostra una predizione del probabile ritardo.

Questo sistema non solo potrà far risparmiare del tempo a chi organizza dei viaggi valutando ogni singola tratta in termini di stazioni e orari di partenza e di arrivo, ma garantirà un risparmio economico ai viaggiatori garantendo che la tratta scelta dal sistema sia la minima e necessaria per arrivare alla destinazione.

1.3 Tool utilizzati

Per la sperimentazione sono stati usati diversi strumenti/librerie, quali:

- **PySWIP**, libreria Python che fornisce un'interfaccia per utilizzare SWI-Prolog, usato per la rappresentazione formale della schedul dei treni
- **NetworkX**, libreria Python utilizzata per la creazione, l'analisi e la manipolazione di reti complesse. Questa libreria fornisce un insieme di strumenti per la rappresentazione di reti e grafi, oltre a un'ampia gamma di algoritmi e funzioni per eseguire diverse operazioni su di essi.

2 Rappresentazione formale della conoscenza

La parte iniziale del progetto si è concentrata sulla rappresentazione formale attraverso fatti e regole Prolog della schedul dei treni e delle stazioni, in particolare ogni treno è caratterizzato da:

1. ID treno, identificatore univoco del treno
2. Tipo di treno, regionale o nazionale
3. ID stazione di partenza

4. ID stazione di arrivo
5. Orario di partenza (HH:MM)
6. Orario di arrivo (HH:MM)
7. Lista delle fermate

Esempio:

```
train(320, nazionale, s01700, s01301, "15:10", "15:58", [s01700, s01307, s01301]).
train(321, nazionale, s01301, s01700, "18:02", "18:50", [s01301, s01307, s01700]).
train(322, nazionale, s01700, s01301, "17:10", "17:58", [s01700, s01307, s01301]).
train(323, nazionale, s01301, s01700, "20:02", "20:50", [s01301, s01307, s01700]).
train(324, nazionale, s01700, s01301, "19:10", "19:58", [s01700, s01307, s01301]).
train(325, nazionale, s01301, s01700, "22:02", "22:50", [s01301, s01307, s01700]).
```

Mentre ogni stazione è caratterizzata da:

1. ID stazione, identificatore univoco delle stazioni
2. Nome stazione
3. Regione stazione

Esempio:

```
station(s11504, "ACQUAVIVA DELLE FONTI", "Puglia").
station(s12026, "ACQUEDOLCI-S.FRATELLO", "Sicilia").
station(s00867, "ACQUI TERME", "Piemonte").
station(s11907, "ACRI BISIGNANO LUZZI", "Calabria").
station(s05420, "ADRIA", "Veneto").
```

Tutte le informazioni relative

3 Analisi del dataset

3.1 Descrizione features

3.2 Preparazione dati

3.2.1 Analisi delle input features

3.2.2 Analisi della target feature

4 Machine Learning

5 Valutazione dei modelli

5.1 Metriche scelte

6 Risultati

6.0.1 Considerazioni

7 Sviluppi futuri

Il sistema presentato è aperto a sviluppi futuri che possano rendere il sistema ancora più efficiente e all'avanguardia.

Di seguito sono descritti alcuni dei possibili sviluppi futuri che intendiamo esplorare:

1. Espansione della copertura ferroviaria andando a completare le informazioni relative a treni e stazioni ed integrazione con altri servizi ferroviari (Italo, Frecciarossa, ...)
2. Implementazione di una vera e propria interfaccia grafica
3. Miglioramento della ricerca andando a suggerire all'utente, sulla base dei caratteri inseriti, le stazioni che iniziano con quei caratteri
4. Integrazione di dati in tempo reale andando a fornire all'utente informazioni in *real time*