

Progetto
Data & Web mining

Zillow Prize: Zillow's Home Value Prediction (Zestimate)

Report Progetto

Prodotto da

Matricola	Nome studente
878169	Alessandro Gasparello
888888	Alessandro Simonato



Università
Ca' Foscari
Venezia

DIPARTIMENTO DI SCIENZE AMBIENTALI ED INFORMATICA

2021/2022

Contents

1	Elaborazione progetto	1
1.1	Introduzione	1
1.2	Creazione del dataframe	1
1.3	Pulizia dei dati e feature engeneering	1
1.4	Analisi dataframe post lavorazione dati	2
1.5	Creazione modelli di regressione e confronti	2
1.6	Conclusioni	2

Chapter 1

Elaborazione progetto

1.1 Introduzione

Nel seguente documento verranno descritte ed illustrate le fasi di sviluppo del progetto **Zillow Prize: Zillow's Home Value Prediction (Zestimate)**. Nello specifico saranno evidenziate e spiegate le decisioni prese, che hanno portato al completamento dello stesso. Saranno utilizzati diversi grafici per mostrare alcuni dei dati importanti che hanno portato alle decisioni finali prese.

1.2 Creazione del dataframe

Inizialmente è stato fatto l'import delle diverse librerie e moduli che sono stati utilizzati all'interno del progetto. Successivamente sono stati caricati i dataset e trasformati in data frame con l'utilizzo della libreria Pandas. Sempre grazie a tale libreria è stato poi fatta l'unione tra i due dataset contenenti diverse feature utili, quali **'logerror'** (la variabile risposta) o la **'transactiondate'** (data acquisto), così da avere tutte le istanze all'interno di un unico data frame.

1.3 Pulizia dei dati e feature engineering

Stampando il contenuto e il tipo delle varie feature contenute nel data frame appena costruito, è risultato che per la maggior parte erano **float64** mentre solo alcune erano **object**. Le prime non presentavano problemi nella gestione, mentre per quanto riguarda le seconde è stato necessario effettuare altre operazioni che si vedranno successivamente per poterle gestire correttamente.

La prima feature che si decise di gestire

1.4 Analisi dataframe post lavorazione dati

1.5 Creazione modelli di regressione e confronti

1.6 Conclusioni