

UNIVERSITA' DEGLI STUDI DI TORINO

MASTER DI I LIVELLO IN ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE

ANNO ACCADEMICO 2019/2020

Titolo tesi: Tecniche di Machine Learning per la predizione della difettosità di un software

Autore: Edoardo Gazzella

Abstract

L'obiettivo del presente progetto è la costruzione di un modello statistico in grado di predire la difettosità di un software. Il progetto vuole essere il prototipo di un prodotto commerciale da proporre ad aziende nel settore ICT come supporto alle decisioni e pianificazione interna.

La difettosità di un software può essere espressa come numero di bug in quattro classi di gravità: Critical, Major, Minor e Trivial. Questa suddivisione necessitava la costruzione di quattro diversi modelli di regressione.

Dopo un periodo di formazione sugli strumenti informatici da utilizzare, a partire dall'obiettivo di predire il numero di bug si sono individuate le variabili indipendenti e quindi le fonti dati necessarie.

Dopo aver raccolto i dati tramite procedure di web-scraping con Python, è stato svolto un attento lavoro di pulizia e normalizzazione, passo fondamentale per la riuscita del machine learning.

Ottenuto il dataset finale, si è deciso di eseguire con il software R diverse tecniche di machine learning, tra cui l'algoritmo Random Forest, con l'obiettivo di ottimizzare quanto più possibile la capacità predittiva del modello.

In seguito, è stata sviluppata una dashboard con il software Qlik Sense per l'analisi descrittiva dei dati e l'analisi predittiva dei risultati.