

Master Universitario in: "ANALISI DATI PER LA BUSINESS INTELLIGENCE"

A.A. 2015-2016

Titolo della tesi: "Previsione della CO2 nel mondo automotive"

Autore: Caterina Barchielli

Abstract

Scopo del progetto di tirocinio è stato quello di costruire un POC di applicazione di tecniche di machine learning in ambito aziendale. In particolare il contesto di applicazione è quello del settore product planning di una casa di costruzione di autovetture.

La pianificazione della produzione di autovetture per il quinquennio a venire deve tener conto dei limiti imposti dalla commissione europea per le emissioni di CO2 previste per i veicoli immatricolati. Il processo di pianificazione prevede che gli ingegneri creino una serie di casi in cui vengono virtualmente assemblate diverse versioni di auto e pianificati i relativi volumi di vendita. Ogni versione di auto viene configurata selezionando le tecnologie opportune per la creazione di motore e impianto di trasmissione. In questa fase è fondamentale trovare il giusto compromesso tra i costi relativi ad alcune scelte tecnologiche per i nuovi modelli e i benefici derivanti dalle ridotte emissioni di CO2.

La CO2 emessa dalle vetture dipende infatti dalla configurazione scelta per assemblare l'autovettura. Per versioni di auto già prodotte tale informazione è già nota e tabulata. Per le nuove versioni di auto la CO2 emessa viene proposta dalla pianificazione e successivamente calcolata e/o misurata dai reparti preposti. L'obiettivo individuato per il POC è stato quindi quello di costruire un modello predittivo che permettesse in fase di pianificazione di avere un'indicazione della CO2 emessa, sulla base della configurazione scelta per la nuova vettura.

Il dataset su cui è stato costruito il modello conteneva i valori di CO2 tabulati per tutte le configurazioni meccaniche principali note. Obiettivo era quindi costruire un modello di regressione che prevedesse il valore di CO2 in base ad una serie di variabili sia categoriche (come ad esempio il tipo di carburante) che continue (come ad esempio la potenza del motore).

La fase di esplorazione e pre-processing del dataset è stata implementata in linguaggio R.

La fase di creazione e confronto dei modelli creati è stata fatta con il supporto di un tool cloud messo a disposizione in versione free da Microsoft: Azure Machine Learning. La scelta dell'algoritmo di regressione più adeguato e dei parametri ottimali è stata fatta tramite cross-validation.

Tramite lo stesso tool è stato poi creato un servizio web che permettesse di richiamare il modello finale creato da un file Excel contenente nuovi dati di input e di restituire i valori predetti.