

Master Universitario in: "ANALISI DATI PER LA BUSINESS INTELLIGENCE"
A.A. 2021/2022

Titolo della tesi: Strumenti e algoritmi per la classificazione di documenti testuali,
con applicazioni al dominio assicurativo

Autore: Elena Veronesi

Abstract

Il presente lavoro ha come obiettivo lo studio di algoritmi di Intelligenza Artificiale, nello specifico di Natural Language Processing, per la classificazione documentale, applicati all'ambiente assicurativo. La normativa vigente richiede la redazione e il continuo aggiornamento di svariate tipologie di documenti, accessori informativi necessari all'emissione di prodotti assicurativi, i così detti documenti PRIIPs.

Questo elaborato è nato da un'attività di tirocinio presso Fairmat srl, un'azienda di Verona, la quale si occupa della trattazione di documenti PRIIPs. Fairmat ha l'esigenza di un processo veloce e accurato che permetta l'identificazione del tipo di documento per poi, successivamente, estrarre specifiche informazioni dal documento stesso per popolare il database OpenKID Vita. Le informazioni da estrarre cambiano in base alla tipologia del documento, pertanto, si è reso necessario identificare in maniera accurata il tipo di documento. L'obiettivo del tirocinio è quello di sviluppare un classificatore, implementato in Python, che permetta di categorizzare il documento secondo diverse tipologie.

Nell'elaborato è quindi introdotto il Natural Language Processing (NLP), sono illustrate le elaborazioni iniziali che si applicano ai dati non strutturati prima di procedere con i modelli e gli algoritmi approfonditi durante il tirocinio. La seconda parte descrive il processo che ha portato all'individuazione del classificatore appropriato per la necessità dell'azienda e riporta i risultati della sperimentazione. Per concludere, vengono presentati brevemente i possibili sviluppi futuri.