Simply Business

Financial Ombudsman Service decision analysis: using public data to understand customer complaints

January 2024

Overview

The Financial Ombudsman Service (FOS) is an institution of critical importance for the insurance business as it helps resolve disputes between consumers and service providers.

When a consumer has a complaint against a financial institution that they have been unable to resolve directly with the company they can bring it to the FOS. The service then examines the evidence from both parties, considering legal principles and regulations and makes a decision on whether the firm is at fault and what the compensation for a customer should be.

Being able to better understand the FOS decisions dataset is of crucial importance for us, as it could help:

- Guide decision making on which products to create,
- Identify potential risks with already existing products,
- Handling of existing complaints.

The goal of this project is to leverage insights obtained from it to minimise future disputes, enhance customer satisfaction, and proactively increase the number and quality of our products and services.

Goals

1. Data Cleaning and Preparation: As the dataset consists of semi-structured text documents it will be necessary to extract the relevant information from the downloaded documents.
2. Natural Language Processing for Data Extraction: Textual data needs to be further analysed to extract relevant information from each document, the data extracted can be used as features for a downstream modelling task.
3. Exploratory Data Analysis: Apply a range of statistical and other modelling techniques to uncover patterns, trends and correlations within the dataset.

4. Visualisation of Trends: This involves visualisation of data distributions, outlier identification, and other interesting insights coming from the dataset.
5. Use of Embedding Techniques: Leverage embedding techniques to transform text data into vector space. This can be particularly useful when it comes to more advanced analysis.
6. Prediction of Claim Outcome: Build and compare at least two binary classification models using the previously extracted features to predict the ultimate FOS decision (Upheld or Not Upheld). Detail how the models perform using classification/misclassification metrics such as recall, precision, F1 score and confusion matrices. Remember to assess the practical usage of your model using information such as feature importance and explainability.

Datasets

The main source of data for this project is the Financial Ombudsman Service Decisions Dataset, a comprehensive collection of records detailing the outcomes of dispute resolutions given by the FOS. This dataset is publicly available but not ready to be used for the tasks proposed here.

We will provide code to download the decisions from the website, it will involve some web scraping and automatic download of PDF files.

Reference

1. Ombudsman decisions dataset: https://www.financial-ombudsman.org.uk/decisions-case-studies/ombudsman-decisions
2. A Practical Introduction to Web Scraping in Python: https://realpython.com/python-web-scraping-practical-introduction/
3. pypdf a Python package to process PDF files
4. SpaCy https://spacy.io/ useful to extract features from the textual data via pattern matching
5. Information Extraction on Tourism Domain using SpaCy and BERT
6. Sentence-Transformers: https://www.sbert.net/ for embedding generation