Machine Learning Course Project Proposal

Software Requirements Classification

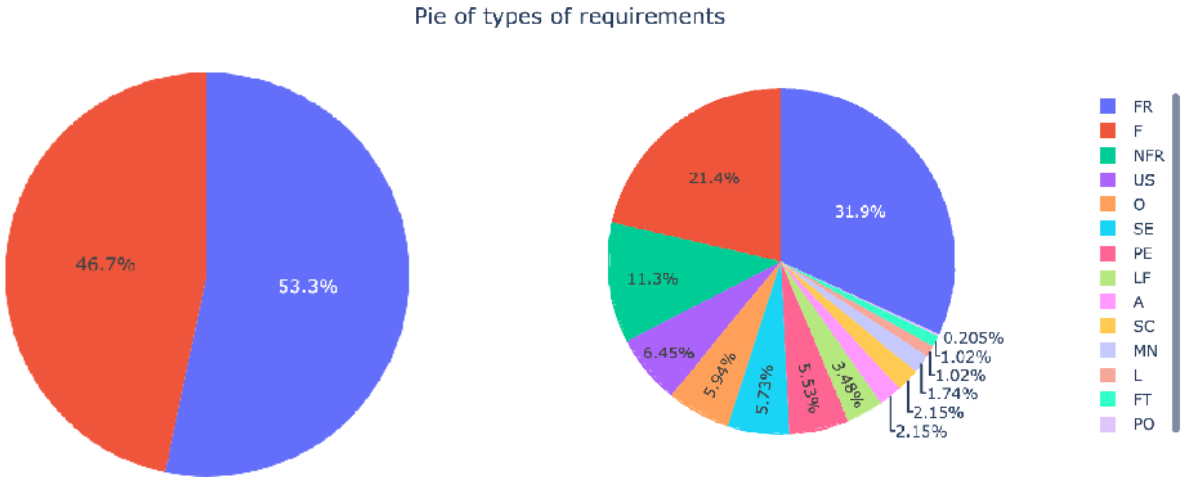
Author: Nikita Breslavsky - 332363498

Introduction

The inception of software projects critically depends on the accurate gathering and classification of requirements. This project aims to leverage machine learning to automatically distinguish between functional and non-functional software requirements, streamlining project management and enhancing development efficiency.

Data

An exhaustive search yielded a singular dataset [1] comprising 1,000 entries of categorized software requirements, initially spread across 14 distinct classes. Given the dataset's limited volume, a granular classification into 14 categories might impede the effectiveness of the project. To circumvent this, a strategic decision was made to combine these into two broad classes: functional and non-functional. This simplification not only diminishes the complexity of the classification task but also promotes a balanced distribution of data. The dataset will be partitioned into a training set (70%) and a test set (30%), facilitating a rigorous model training and evaluation process.



Machine Learning Models:

Employing the Scikit-learn library, the project will develop, fine-tune, and evaluate the following models:

- Logistic Regression
- Support Vector Classification (SVC)
- K-Nearest Neighbors (KNN)
- Naive Bayes
- Decision Trees (with and without AdaBoost)
- Voting Classifier (combining the above models)

Each model's performance will be rigorously compared to identify the most efficient algorithm for our classification task.

Bibliography

[1] "Software requirements dataset," [Online]. Available: https://www.kaggle.com/datasets/iamvaibhav100/software-requirements-dataset.