NLP Grievances Classification Project

Introduction

This project involves an NLP-based classification model built on a grievances dataset. The aim is to analyze textual grievance data and build a model that can classify or predict issues from unstructured complaints.

Dataset Overview

The dataset is loaded and consists of several hundred thousand rows. It includes grievance texts and possibly corresponding labels or metadata for classification tasks. Data sampling and slicing is performed to limit the dataset for practical training.

Data Preprocessing

Text data is cleaned and processed using techniques like lowercasing, punctuation removal, stopword filtering, and lemmatization to make it suitable for vectorization and model training.

Feature Engineering and Vectorization

TF-IDF vectorization is used to convert the cleaned text into numerical features suitable for model input. This helps capture the importance of words across documents.

Model Building

Several machine learning models such as Logistic Regression, Random Forest, or Naive Bayes may have been used to classify the grievances. Model performance is evaluated using accuracy, confusion matrix, and classification report.

NLP Grievances Classification Project

Conclusion

The model successfully demonstrates the use of NLP techniques for text classification tasks. This approach is scalable and applicable to similar problem statements in industries like customer service, insurance, or telecom.