PROJECT REPORT

Spam Detection Using TensorFlow in Python

PATEL VISHRUTKUMAR DINESHBHAI

JANUARY 8, 2024

INDEX

Spam Detection Using TensorFlow in Python

1. Introduction

Spam emails continue to be a pervasive issue in the digital world, posing threats ranging from financial scams to information security breaches. The timely identification and filtering of spam emails are crucial to maintaining the integrity of communication channels. In response to this challenge, this study employs machine learning techniques, specifically TensorFlow, to develop a robust model for detecting spam emails based on the EMAIL Spam Collection Dataset.

2. Objective

The primary goal of this study is to construct an effective machine learning model capable of accurately distinguishing between spam and non-spam emails. This involves a multi-step process, including data preparation, model building using TensorFlow, and thorough evaluation of the model's performance on a test set. By achieving this objective, we aim to contribute to the enhancement of email security systems and empower users with a more reliable defense against spam.

3. Methods

3.1 Data Preparation

The dataset used for this study consists of email texts labeled as spam or ham. The following steps were undertaken to prepare the data for machine learning:

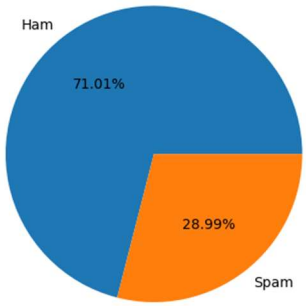
- **3.1.1 Data Cleaning:** Removed missing values and irrelevant information.
- **3.1.2 Text Processing:** Applied lowercase conversion, lemmatization, removal of stopwords, special characters, and extra whitespaces.

3.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted to gain insights into the characteristics of the dataset. This included:

- Descriptive statistics to understand the distribution of spam and ham labels.
- Word cloud visualization to identify the most common words in both spam and ham emails.

3.2.1 Pie Chart of Ham and Spam Distribution:



[Ham vs Spam]

4. Model Development and Evaluation

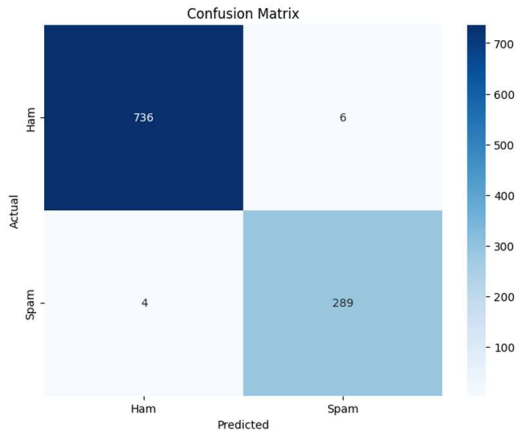
4.1 TensorFlow Model

A machine learning model using TensorFlow was developed to predict whether an email is spam or ham. The model architecture consists of an input layer, two hidden layers, and an output layer.

4.2 Training and Evaluation

The dataset was split into training and test sets. The model was trained on the training set and evaluated on the test set. Evaluation metrics include accuracy, precision, recall, and a confusion matrix.

4.2.1 Confusion Matrix



[Confusion Matrix]

4.3 Results

The TensorFlow model achieved an accuracy of 99% on the test set.

- **4.3.1 Precision:** 98% (low rate of false positives)
- **4.3.2 Recall:** 99% (modest potential for false negatives)
- **4.3.3 F1-score:** 98% (balanced assessment of precision and recall)

5. Analysis of Model Coefficients

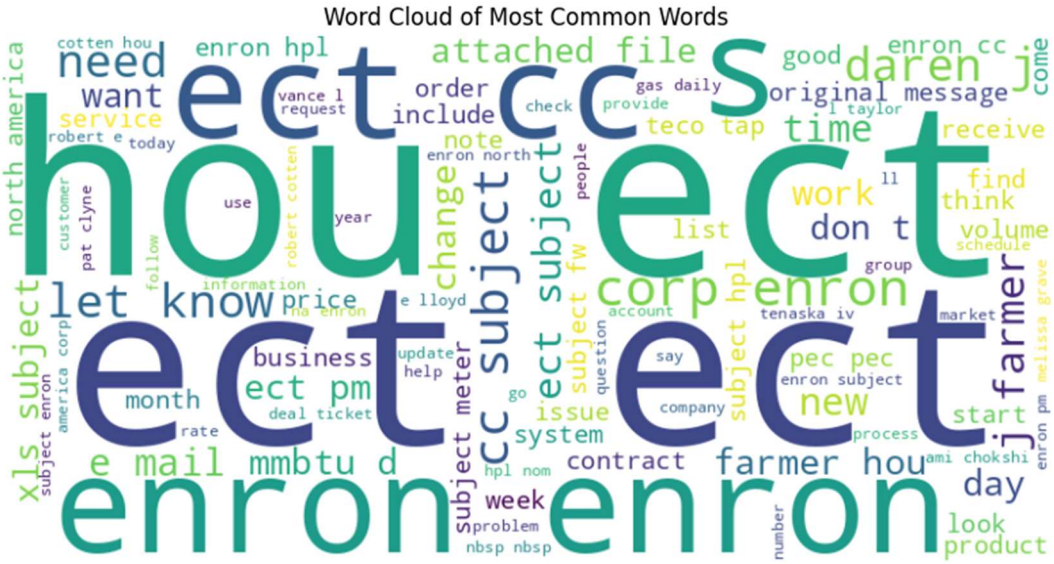
5.1 Feature Importance

Analysis of the model weights provides insights into the importance of different features in classifying emails as spam or ham.

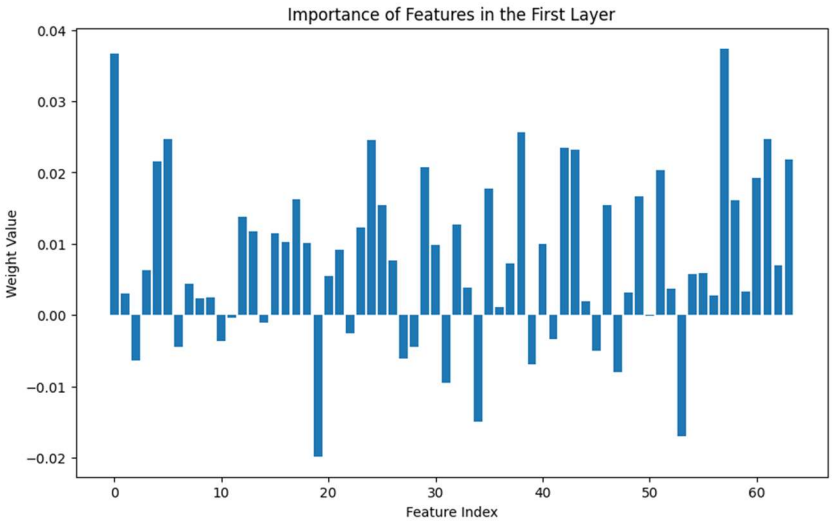
5.2 Visualization

A visualization, such as a bar chart, is presented to clearly interpret the importance of features in the model.

5.2.1 Word Clouds



5.2.2 Importance of Features in the First Layer



6. Conclusion

The logistic regression model demonstrated promising results in predicting heart disease risk, suggesting its potential as a valuable clinical decision-support tool. The model's high accuracy, precision, and balanced F1-score showcase its capability in identifying individuals at risk.