Intelligent Email Classifier Using Machine Learning

This project focuses on building a machine learning model to classify text messages as spam or not spam. With the rise of digital communication, unsolicited and fraudulent messages have become a major concern. Accurate spam detection helps protect users from scams, phishing attempts, and unnecessary interruptions.

Objective

The primary objective of this project was to develop a reliable text classification model that can automatically distinguish between spam and legitimate (ham) messages. This supports improved user experience and enhances security in communication platforms.

Methodology

1. Data Collection: Utilized a publicly available SMS spam dataset containing labeled messages.
2. Data Cleaning: Removed duplicates, handled missing values, and normalized text data by lowering case, removing punctuation, and eliminating stop words.
3. Feature Engineering: Transformed text data into numerical form using TF-IDF and CountVectorizer.
4. Model Development: Trained multiple models including Logistic Regression, Naive Bayes, and Random Forest.
5. Evaluation: Compared models using Accuracy, Precision, Recall, F1-score, and ROC-AUC
6. Model Deployment: It was deployed using hugging face.

Results & Impact

- Naive Bayes performed best for this task due to its strength in handling text classification, achieving high precision and recall.
- The model demonstrated over 95% accuracy, minimizing both false positives (ham misclassified as spam) and false negatives (spam misclassified as ham).
- This system can be integrated into communication services to filter unwanted messages and enhance user trust and safety.

Link (Optional)

https://huggingface.co/spaces/ebytt12/Intelliegent_Email_Classsifier

Relevance to Data Science MSE

The spam detection project highlights the application of natural language processing (NLP), feature engineering, and classification algorithms in real-world problems. It demonstrates how data science can be leveraged to improve security, user experience, and operational

efficiency. The project is closely aligned with the MSE program's focus on practical applications of machine learning and model evaluation.