

Final Project Statement

InboxIQ–Intelligent Email Filtering with ML

Winter in Data Science (WiDS)

December 30, 2024

Project Overview

The InboxIQ project is designed to introduce participants to deep learning concepts for email classification. You will learn how to preprocess data and implement a Recurrent Neural Network (RNN) for spam/ham email classification. Our project is divided into two parts for better understanding and execution.

Learning Objectives

1. Understand how to preprocess email datasets for classification.
2. Implement an RNN model using TensorFlow.
3. Learn the concepts of epochs, early stopping, and validation splits.
4. Evaluate model performance using appropriate metrics.

Project Details

Part 1: Data Preprocessing

In this part, you will:

1. Import the dataset (`spam_ham_dataset.csv`).
2. Read and clean the data as needed.
3. Tokenize and vectorize the email text for model input.

Part 2: RNN Implementation

In this part, you will:

1. Split the data into 80% training and 20% validation sets.
2. Build and compile an RNN model using TensorFlow/Keras.
3. Train the model with appropriate hyperparameters.
4. Use early stopping to prevent overfitting.
5. Evaluate the model's performance.

Submission Instructions

- Submit your code in two parts: Part 1 (data preprocessing) and Part 2 (model implementation).
- Add comments for any additional steps you take.
- Submit your completed project on google forms.
- Submissions received after **January 7, 11:59 PM** will not be considered for certification.

Your Mentors:

Feel free to reach out to us with any questions or concerns:

- **Yaswanth:** 7989068522
- **Avyay:** 9550104004

Thank You!