

Report

Internship Code for Building a Spam Detection Model using TensorFlow and Keras This internship code confidently creates a binary text classification model using TensorFlow and Keras to differentiate between spam and non-spam (ham) messages. The primary objective of this project is to establish a solid foundation for a more robust spam detection system. The code begins by importing essential libraries such as TensorFlow, Pandas, NumPy, matplotlib, and scikit-learn. These libraries are used for various functions, including data manipulation, model building, and evaluation. Next, the dataset is loaded from a CSV file named "spam_ham_dataset.csv" using Pandas.

Basic information about the dataset is displayed using `df.info()`, and the first few rows of the dataset are shown with `df.head()`. Unnecessary columns ('Unnamed: 0' and 'label') are dropped from the DataFrame. The input features (x) and labels (y) are separated from the Data Frame, and the dataset is split into training and testing sets using `train_test_split()`. Tokenization and Padding are then applied to the training text data (X_train) using Tokenizer from Keras. Padding ensures a uniform length of sequences using pad sequences.

Next, a Sequential model is created using Keras, consisting of an Embedding layer for word embeddings, a Flatten layer to flatten the input, followed by Dense layers with ReLU and sigmoid activations for classification. The model is compiled using the Adam optimizer and binary cross-entropy loss function. The model is then trained using `model.fit()` on the padded training data (X_train_pad) and corresponding labels (Y_train) for 5 epochs with a validation split of 0.2.

Predictions are made on the test data (x_test_pad) using `model.predict()`. Predictions are thresholded at 0.5 to obtain binary predictions. Finally, the binary predictions are displayed. This code confidently implements a basic text classification model using TensorFlow and Keras. With further refinement and evaluation, this code has the potential to serve as a strong foundation for building more robust spam detection systems.