Progress Report

16/08/2022

Overview

- Using datasets of sms messages labeled `spam` if they are considered spam, and `ham` if considered otherwise
- Generating a model capable of classifying incoming messages to the above categories
- Building a Telegram bot using `Pyrogram` package, to use the trained model
- Deploying the ML model to a Telegram groupchat via the bot

EDA

Taks performed include:

- Concatenating various datasets to increase my inventory of data
- Exploring the distribution of the messages, such as how long are they in word length, for both `ham` and `spam` labels
- Fixing imbalance problem through `over-sampling`

Next Phases

Phase	Date
Text Vectorization	17/08/2022
Model Development	18/08/2022
Model Evaluation	18/08/2022
Creating python script for Telegram bot	19/08/2022
Implementing the ML model in the bot script	20/08/2022
Model Deployment	21/08/2022
Preapraing slides and repo for submission	22/08/2022