Assignment - 4

SMS Spam Classification

|  |  |
| --- | --- |
| Assignment Date | 08 October 2022 |
| Student Name | KHAMILA BANU K |
| Student Roll Number | 2019504539 |
| Maximum Marks | 2 Marks |

Problem Statement:

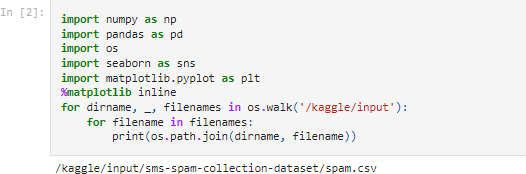
Problem Statement:

Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollar industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. Due to Spam SMS, Mobile service providers suffer from some sort of financial problems as well as it reduces calling time for users. Unfortunately, if the user accesses such Spam SMS they may face the problem of virus or malware. When SMS arrives at mobile it will disturb mobile user privacy and concentration. It may lead to frustration for the user. So Spam SMS is one of the major issues in the wireless communication world and it grows day by day.

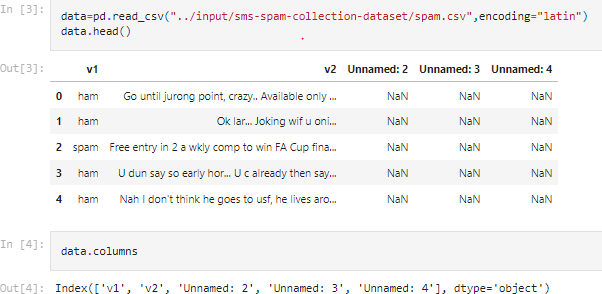
Solution: Spam Message Classification using LSTM

Source code and corresponding outputs:

1. Import the Necessary Libraries

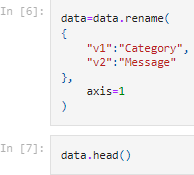


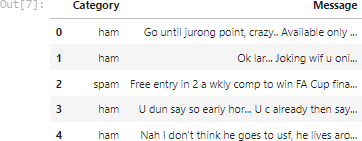
1. **Reading the .csv dataset**



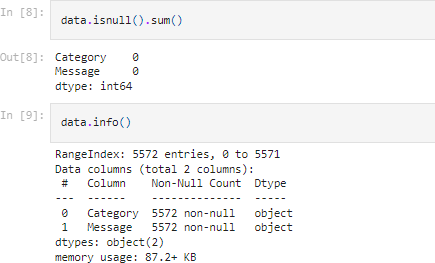
1. **Drop the unnamed Columns**



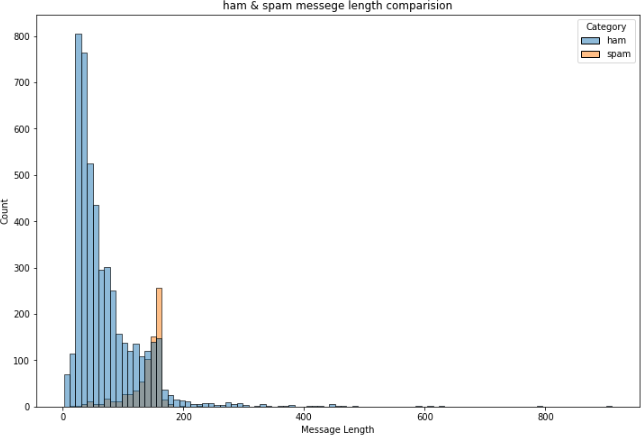
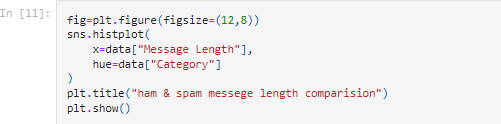
1. **Renaming Column names sensible**

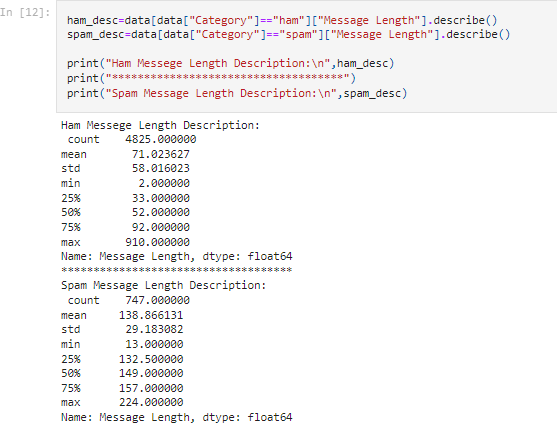


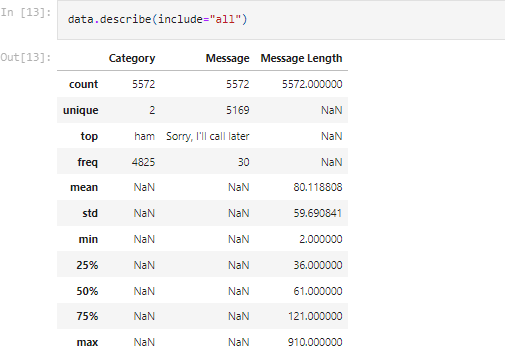
1. **Check for null values in dataset**



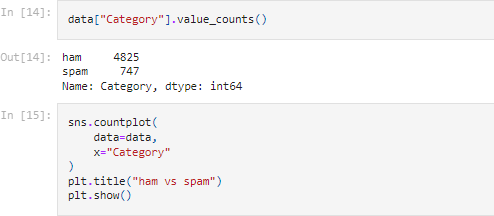
1. **Creating a new Field to store the Message Lengths**
2. **Histogram Inference of Message Lengths of Spam and Non-spam messages**

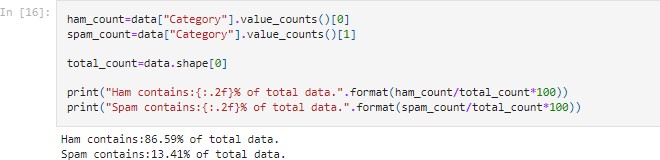
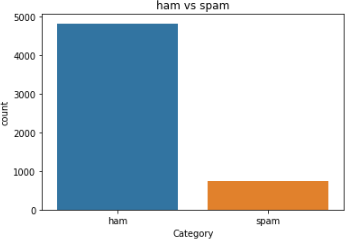






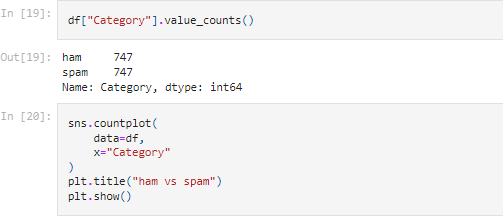
1. **VISUALIZING COUNT OF MESSAGES OF SPAM AND NON SPAM**

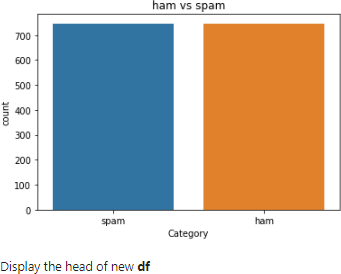


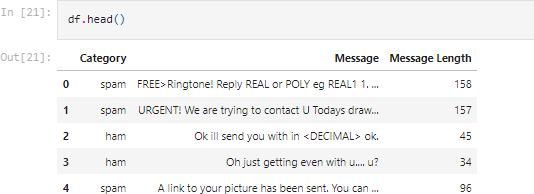


1. **Undersampling to Genralize Model and Balance Spam and Ham quantities in dataset**

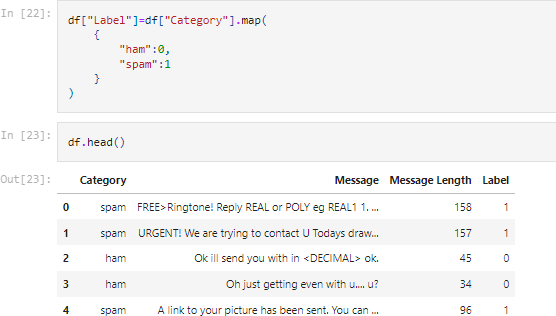








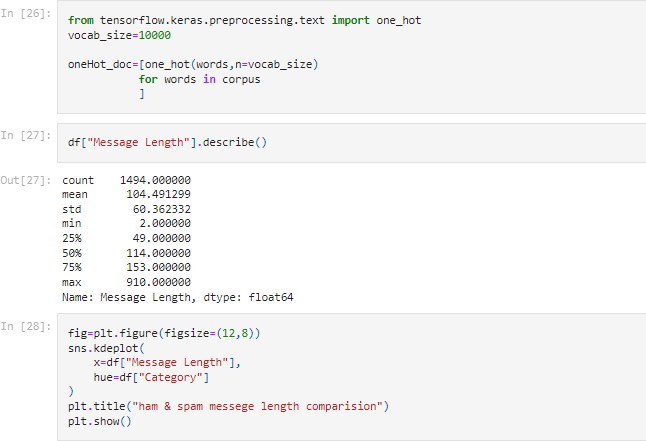
1. **Binary Encoding of Spam and Ham Categories**

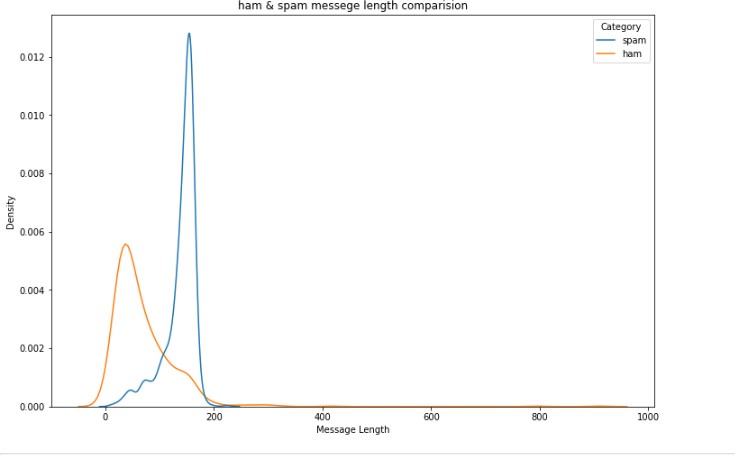


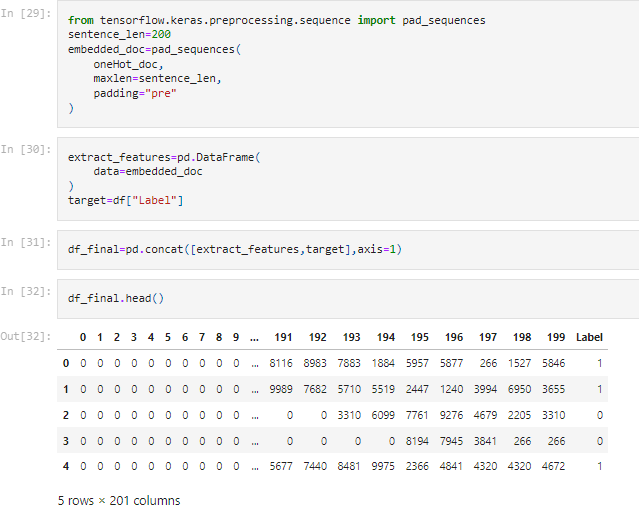
1. **Import Necessary Libraries to perform Word Tokenization**



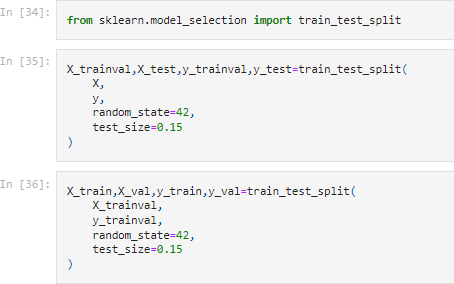
1. **Perform One Hot on Corpus**



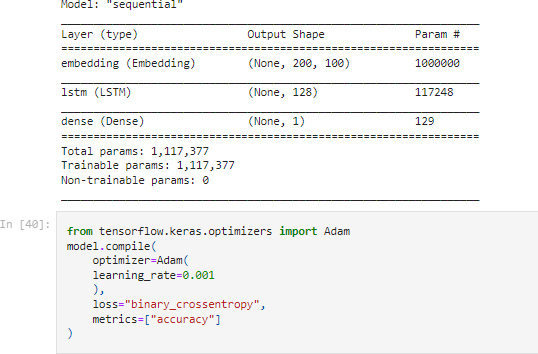




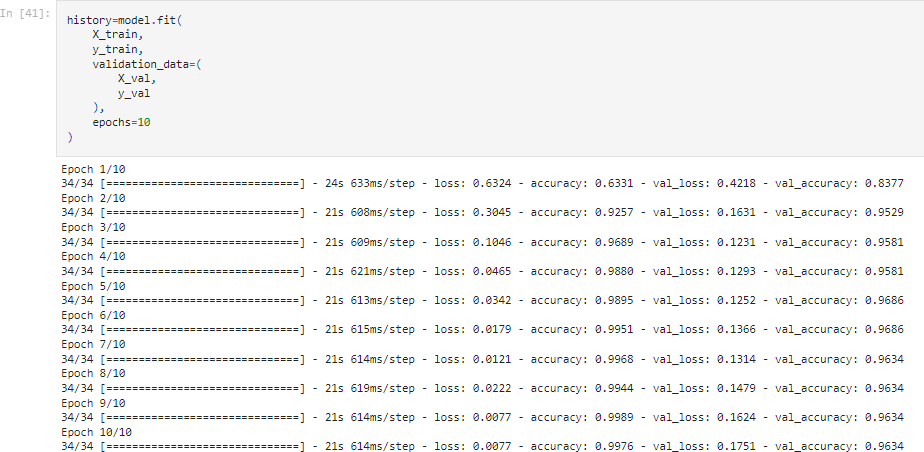
1. **Splitting Dependent and Independent Variables**
2. **Train, test and Validation Split**



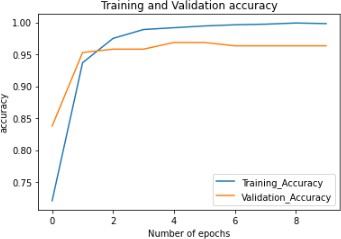
1. **Building a Sequential Model**

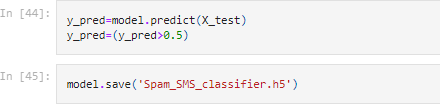


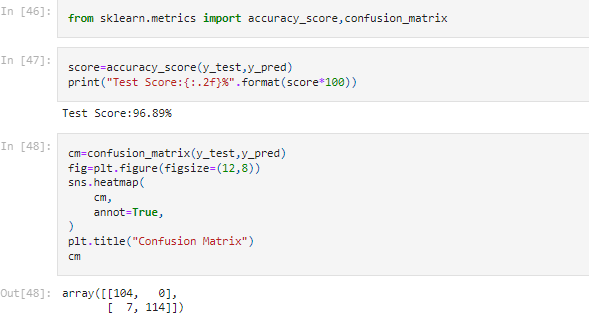
1. **Model Fitting**

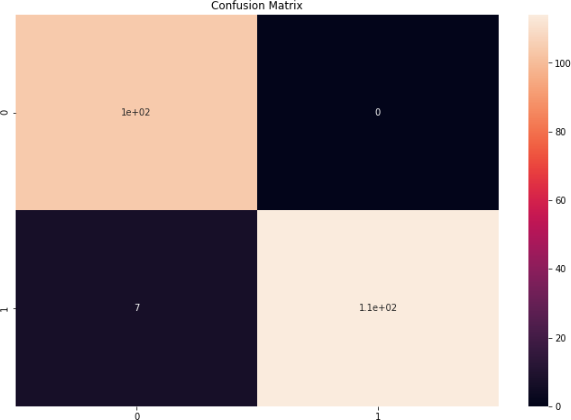






1. **Save and Test the Model**
2. **Evaluating the Model**





1. **Function to Test the Model on a Random message**

