**Date of Submission: 31st March,2022**



Project Proposal

CS - 4051

**Sarcasm Detection in News Headlines**

**Using SVM**

**Group Members:**

1. Muhammad Zunique (20K-0145)
2. Sanan Baig (20K-0165)
3. Awwab Sabir (20K-1615)

**Course Instructor:** Dr. Muhammad Rafi

**INTRODUCTION:**

The purpose of this project is to develop a machine learning model that can accurately detect sarcasm in news headlines using SVM. This model can be used to automatically identify satirical news articles or clickbait headlines, which are often written in a sarcastic or humorous tone, from legitimate news articles. This is important because the spread of fake news and misinformation can have serious consequences for individuals and society as a whole. By detecting sarcasm in news headlines, this model could help people make more informed decisions and reduce the impact of false information.

**What is SVM?**

**SVM** stands for Support Vector Machines. It is a machine learning algorithm that is commonly used for classification and regression analysis.

It is a technique that helps us draw a boundary between different types of data. The boundary is chosen to have the most space possible between the closest points of each group. This method works well for all kinds of data, even when there are many different factors to consider. **SVMs** are powerful algorithms that can handle both linear and nonlinear data, and they are effective even in high-dimensional spaces. They are widely used in applications such as image classification, text classification

**Dataset:**

The dataset used for this project model will be news headline dataset provided by ***Dr. Muhammad Rafi*** which contains **26,000+** news headlines over various sources. The Headlines Dataset consists of three attributes: Article link: link to the original news article Headline: the headline of the news article is\_sarcastic: 1=sarcastic; 0=non-sarcastic.

**Approach:**

The approach to the project model will be as follows:

1. **Data Pre-processing:**

Cleaning of dataset will be performed in this step, by removing stopwords, punctuation and other special characters and by case folding it into lowercase/uppercase by using some python Libraries.

1. **Features Extraction:**

In this step, we will extract some important features from our cleaned data that would help to predict our model accuracy, such as n-grams and word embedding.

1. **Model Development:**

The next step will be to train and evaluate a Support Vector Machine (SVM) model using the extracted features. SVM is a powerful classification algorithm that has been used successfully for various text and image classification.

1. **Model Evaluation:**

The performance of the SVM model will be evaluated on various performance metrics, such as accuracy, precision, recall, and F1-score. The model will also be tested on a separate test dataset to evaluate its generalization performance.

**Expected Outcome:**

The goal of this project is to create a Support Vector Machine (SVM) model that can identify sarcasm in news headlines with high accuracy. The model will be assessed using different performance metrics and tested on a separate dataset to determine how well it can perform on new data.

**Conclusion:**

This project aims to develop a machine learning model that can accurately detect sarcasm in news headlines, which can be used to identify satirical or misleading news articles. The outcome of this project can have significant implications for reducing the spread of fake news and misinformation in the media. The project will be implemented using Python programming language and several machine learning libraries.