TITLE OF THE PROJECT –

The title of the project is **Email Spam Detection** based on Machine Learning. In this project we compare the best ML model by using several algorithms.

OBJECTIVE OF THE PROJECT –

- Understand the Dataset & clean-up (if required).
- Build classification models to predict whether or not the email is spam.
- Also fine-tune the hyperparameters & compare the evaluation metrics of various classification algorithms.

PROJECT DESCRIPTION –

The "spam" concept is diverse: advertisements for products/web sites, make money fast schemes, chain letters, pornography, etc.

Our collection of spam e-mails came from our postmaster and individuals who had filed spam. Our collection of non-spam e-mails came from filed work and personal e-mails, and hence the word 'george' and the area code '650' are indicators of non-spam. These are useful when constructing a personalized spam filter. One would either have to blind such non-spam indicators or get a very wide collection of non-spam to generate a general purpose spam filter.

The dataset, taken from the UCI ML repository, contains about 4600 emails labelled as **spam** or **ham**.

SOFTWARE REQUIREMENTS –

Software requirement for this project are –

➢Operating System (64-bit) – Windows OS (latest version)

➢We need to have the dataset files in a .csv extension.

➢IDEs – Jupyter Notebook/ Google Colab (Preferred) (Install Jupyter Notebook or any similar working environment with the latest version of python installed)

➢We must know the Python language and its libraries which includes NumPy, Pandas, Matplotlib, and scikit learn, etc.

➢Web Browser (At least any one) –

- Mozilla Firefox (latest)
- Google Chrome (latest)
- Internet Explorer 11.x

HARDWARE REQUIREMENTS –

Hardware requirement for this project are –

- ➢ Processor – i3 (minimum)
- ➢ Hard Disk – 1 TB (preferred)
- ➢ Memory – 2GB RAM (minimum)
- ➢ Screen Resolution (Optional) – 1280*1024 (to view the Console)
- ➢ Input Devices: Keyboard, Mouse

ADVANTAGES –

- ❖ Spam detection is proficient of filtering mails giving to the content of the email and not according to the domain names or any other criteria.
- ❖ Good Efficiency
- ❖ Greater accuracy comparison using different ML Algorithms.

DISADVANTAGES –

- ❖ This project in not 100% accurate.
- ❖ It is possible to make mistake.
- ❖ Work differently in different environment.

FUTURE SCOPE –

Review spam detection is essential since it can ensure justice for the sellers and retain the trust of the buyer on the online stores. The algorithms developed so far have not been able to remove the requirement of manual checking of the reviews. Hence there is scope for complete automation of spam detection systems with maximum efficiency. With growing popularity of online stores, the competition also increases. The spammers get smarter day by day and spam reviews become untraceable. It is necessary to identify the spamming techniques in order to produce counter algorithms.

CONCLUSION –

Email has been the most important medium of communication nowadays, through internet connectivity any message can be delivered to all over the world. More than 270 billion emails are exchanged daily, about 57% of these are just spam emails. Spam emails, also known as non-self, are undesired commercial or malicious emails, which affects or hacks personal information like bank, related to money or anything that causes destruction to single individual or a corporation or a group of people. Besides advertising, these may contain links to phishing or malware hosting websites set up to steal confidential information. Spam is a serious issue that is not just annoying to the end-users but also financially damaging and a security risk. Hence this system is designed in such a way that it detects unsolicited and unwanted emails and prevents them hence helping in reducing the spam message which would be of great benefit to individuals as well as to the company. In the future this system can be implemented by using different algorithms and also, more features can be added to the existing system.