SOCIAL / BUSINESS IMPACT

Date	17 March
Team Id	NM2023TMID18550
Project Name	Optimize Spam Filtering
Maximum Marks	5Marks

THIS IS A SMALL IMPACT OF –OPTIMIZE SPAM FILTERING

✓ They may send spam emails that contain attachments with the multiple-file extension, packed URLs that lead the user to malicious and spamming websites and end up with some sort of data or financial fraud and identify theft [4, 5]. Many email providers allow their users to make keywords base rules that automatically filter emails. Still, this approach is not very useful because it is difficult, and users do not want to customize their emails, due to which spammers attack their email accounts.

✓ These approaches have their limitations and drawbacks. There has been significant growth in spam emails, along with the rise of the Internet and communication around the globe [6]. Spams are generated from any location of the world with the Internet's help by hiding the attacker's identity. There are a plenty of antispam tools and techniques, but the spam rate is still very high.

✓ Search engines – Another way to encourage people to access the website of an electronic business is done through a registration on various search engines like Google, Lycos, HotBot or Yahoo. There are search engines that automatically index these websites, with links from other pages to this address.

✓ The use of specializing firm in sending messages, a company that has made a reputation with anti-spam policies formulated and implemented well. In this way, it can be avoided the blocking of these emails to different ISPs, and the rate of messages arriving in inbox, from junk mail, is higher.

It is a stream of data science where natural language data is given to machine learning algorithms for various purposes like: –

- ✓ Sentiment Analysis
- ✓ Spam filtering
- ✓ Named entity recognition
- ✓ Part of speech tagging etc.
- ✓

- ✓ Opinion Rank detects more trustworthy websites and lesser spam websites with a shorter time. One of its disadvantages is, it is very challenging to identify the subset of websites which are needed to update its trust value.

- ✓ It consists of four modules which are preparation of email dataset, pre-processing of data, feature selection and hybrid bagged approach. Only the J48 algorithm gives the experimental results better. Other two experiment gives low performance. To enhance the system's performance by using the concept of boosting approach. It will replace the features of weak classifier learning features with a strong classifier's approach.

- ✓ Using the famous topic modelling technique, Latent Dirichlet Allocation(LDA) topic classification is done by giving the dynamic topics. LDA is the generative statistical model which explains the similarity of data. LDA classifies the document to a specific topic. Then using the java mail API, advertisement mails are detected and deletion operation is performed.

- ✓ Nowaday, emails are used in almost every field, from business to education. Emails have two subcategories, i.e., ham and spam. Email spam, also called junk emails or unwanted emails, is a type of email that can be used to harm any user by wasting his/her time, computing resources, and stealing valuable information.