AN INTERNSHIP REPORT

On Project

SPAM EMAIL/SMS DETECTION

At

COMPOZENT

A report submitted in partial fulfillment of the requirements for the Award of Degree of

BACHELOR OF ENGINEERING

In

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

By

GANESH JAHAGIRDAR (TE 17, AI AND DS 2023-24)

Under Supervision of

Mr. Harshkumar Vishwakarma (Internship Mentor), Compozent and

Prof. Namrata D. Ghuse, KKWIEER

(Duration: 15 Dec 2023 – 15 Jan 2024)

K. K. WAGH INSTITUTE OF ENGINEERING EDUCATION AND RESEARCH, NASHIK, MAHARASHTRA.

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

STUDENT'S DECLARATION

I, **Ganesh Shripad Jahagirdar** hereby declare that I have undertaken 04 weeks internship at **Compozent** during a period from 15th Dec 2023 to 15th Jan 2024 in partial fulfillment of the requirements for the Award of Degree (ARTIFICIAL INTELLIGENCE AND DATA SCIENCE) at **K. K. WAGH INSTITUTE OF ENGINEERING EDUCATION AND RESEARCH, NASHIK.** The work which is being presented in the training report submitted to the Department of ARTIFICIAL INTELLIGENCE AND DATA SCIENCE at above mentioned institute is an authentic record of training work.

I have taken care in all respect to honor the intellectual property right and have acknowledged the contribution of others for using them in academic purpose and further declare that in case of any violation of intellectual property right or copyright I, as a candidate, will be fully responsible for the same.

Signature of the Student
Date:
Place:

Mr. Harshkumar Vishwakarma (Internship Mentor), Compozent
Signature of the Supervisor-1

Prof. Namrata D. Ghuse, KKWIEER
Signature of the Supervisor-2

ACKNOWLEDGEMENT

It is always a pleasure to remind the fine people in the Engineering program for their sincere guidance I received to uphold my practical as well as theoretical skills in engineering.

Successful completion of any type of training requires helps from number of people. Also, the help needed to prepare this report from different people cannot be overlooked. Now there is a little effort to show my gratitude to every person helped in these four weeks program.

Secondly, I want to thank **Harshkumar Vishwakarma (Internship Mentor)** for giving me this opportunity to do an internship/industrial training in the esteemed company.

I would like to convey my heartiest thanks to **Mr. Harshkumar Vishwakarma (**Senior Software Engineer, Technical Supervisor at Compozent) who despite being extraordinarily busy with their duties, took time to hear, guide and keep me on the correct path and allowed me to carry out my training in the company.

I would also like to express my gratitude to **Prof. Dr. S. S. Sane** (Head of Department), **Prof. Namrata D. Ghuse** (Internal Guide) and faculty from the college, for helping me and regularly maintaining the supervision from the college side. And, last but not the least, I express my deepest thanks to all the departments and staff at Compozent.

SUMMARY

An internship is a period of work experience offered by a company/an organization for a limited period of time. It is an opportunity that employers offer to a student interested in gaining work experience in their company. The report presents the work I have done, the knowledge has been acquired and the conclusions I have drawn in these 04 weeks internship/ industrial training at compozent.

This report presents a project on spam email/SMS detection using machine learning. The project aimed to develop a machine learning model that could accurately classify email messages as spam or non-spam. The dataset used for the project consisted of 5,573 emails, with a 13:87 split between spam and non-spam emails, obtained from Kaggle.

The project involved several methodological details, including data Cleaning, EDA, Text Preprocessing, model building and evaluation. The selected model achieved an accuracy of 99% and precision of 98% on the test set.

In conclusion, the project successfully developed a machine learning model for email spam detection that achieved high accuracy. Future work could explore the use of deep learning models for email spam detection. Overall, the project highlights the effectiveness of machine learning algorithms in detecting spam emails and their potential to improve digital communication.

INDEX

Chapter – 1
INTRODUCTION

In today's interconnected world, email communication has evolved into a fundamental aspect of daily life, with millions of emails exchanged regularly. However, the ubiquity of legitimate emails is accompanied by the persistent challenge of spam, which not only causes inconvenience but also poses potential harm to recipients. To combat this issue, the field of email spam detection has become crucial in computer science, employing machine learning techniques to build automated systems capable of distinguishing between spam and non-spam emails. Recognizing the broader scope of text-based communication, this report extends its focus beyond email to include the detection of spam in Short Message Service (SMS) as well. Furthermore, the proposed model is designed to be versatile, suggesting its applicability to various text-based software for spam detection, thus contributing to a more comprehensive approach to cybersecurity.

The report aims to present a detailed account of a project cantered around email and SMS spam detection, utilizing machine learning methodologies. It delves into the specifics of the project, including the chosen dataset, the methodology employed, and the resultant outcomes. Additionally, the report underscores the significance of email and SMS spam detection, outlines the challenges inherent in building effective spam filters, and explores the potential transformative impact of machine learning algorithms on digital communication. Commencing with an overview of the vital role played by email and SMS communication in contemporary society, the report then addresses the persistent challenges posed by spam in these channels. It introduces the concepts of data science and machine learning, elucidating how these techniques can be harnessed to develop robust spam filters. The subsequent sections detail the project, covering the dataset selection, methodology, and results. Finally, the report concludes with a broader discussion on the prospective influence of machine learning algorithms in elevating the quality of digital communication and suggests avenues for future research in this dynamic field.

1.1 What is AI & ML?

a) Artificial Intelligence:

Artificial Intelligence refers to the development of computer systems or software that can perform tasks that typically require human intelligence. These tasks include problem-solving, learning, understanding natural language, recognizing patterns, speech recognition, and decision-making. AI aims to create machines that can mimic cognitive functions such as learning and problem-solving, enabling them to adapt to new situations and perform tasks without explicit programming.

There are two main types of AI:

1.Narrow AI (Weak AI): This type of AI is designed to perform a specific task or a narrow set of tasks. Examples include voice assistants, image recognition systems, and recommendation algorithms.

2.General AI (Strong AI): General AI refers to machines or systems with the ability to perform any intellectual task that a human being can. Achieving true general AI is a complex and ongoing challenge.

b) Machine Learning (ML):

Machine Learning is a subset of AI that focuses on the development of algorithms and statistical models that enable computers to perform tasks without being explicitly programmed. Instead of relying on explicit programming, ML systems learn from data, identifying patterns and making decisions or predictions based on that learning.

Key concepts in Machine Learning include:

1. **Supervised Learning**: The algorithm is trained on a labeled dataset, where the input data is paired with the corresponding desired output. The model learns to map inputs to outputs.

2. **Unsupervised Learning**: The algorithm is given unlabeled data and must find patterns or relationships within the data without explicit guidance. Clustering and dimensionality reduction are common unsupervised learning tasks.

3**. Reinforcement Learning**: The algorithm learns by interacting with an environment and receiving feedback in the form of rewards or penalties. The goal is to learn a policy that maximizes cumulative rewards over time.

Machine Learning can be applied to various domains, such as natural language processing, image recognition, recommendation systems, fraud detection, and more.

1.2 Machine Learning Life Cycle

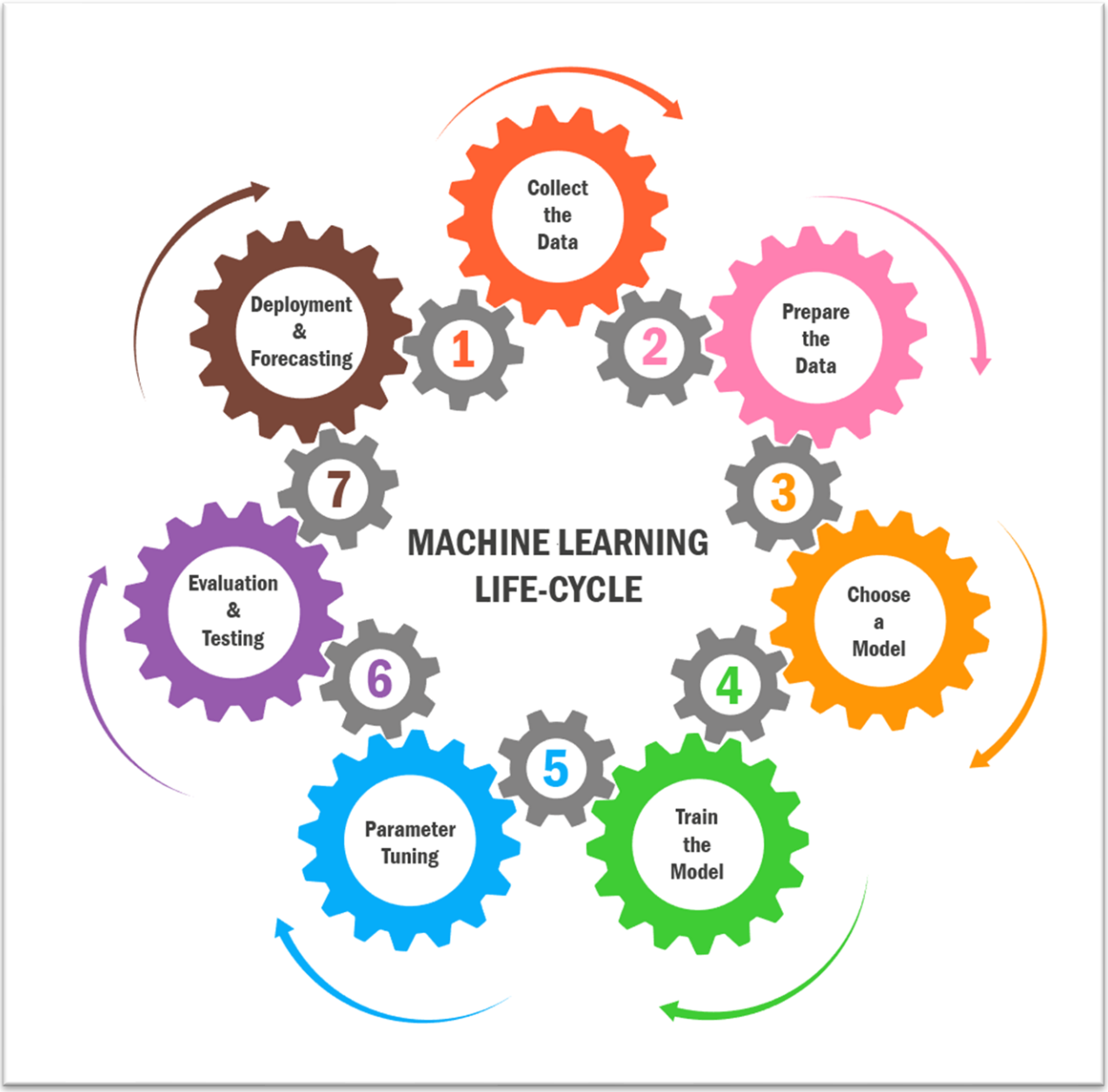

Fig.1.1 Machine Learning Life Cycle

The data science life cycle is a systematic approach to solving complex problems using data-driven techniques. It involves several stages or steps that guide the process from understanding the business problem to presenting insights and findings. Here are the 8 steps in the timeline of Machine Learning Project:

1. Data Cleaning:

Involves identifying and correcting errors or inconsistencies in the dataset, handling missing data, and ensuring the data is in a format suitable for analysis.

2. EDA (Exploratory Data Analysis):

Visually and statistically explores the dataset to understand its main characteristics, identifying patterns, trends, relationships, and potential outliers.

3. Text Preprocessing:

Essential for tasks involving text, such as natural language processing. Includes tasks like tokenization, stemming, lemmatization, and removing stop words to prepare the text for analysis.

4. Model Building:

Involves selecting an appropriate machine learning or statistical model based on the nature of the task (classification, regression, clustering, etc.) and training the model on the dataset to learn patterns and make predictions.

5. Evaluation:

After training the model, assess its performance using metrics such as accuracy, precision, recall, and F1-score, depending on the nature of the problem.

6. Improvement:

Based on the evaluation results, fine-tune the model by adjusting hyperparameters, trying different algorithms, or obtaining additional data to enhance performance.

7. Website:

Creating a user interface or platform for the model, possibly displaying results, providing user input options, or showcasing visualizations.

8. Deploy:

Making the model accessible for others to use, whether by integrating it into a web application, creating an API, or incorporating it into a larger system for real-world applications.

It is important to note that these steps can be iterative, and the process may loop back to earlier steps based on the outcomes of evaluation and improvement. Additionally, communication and collaboration between different steps are crucial for the success of the project. These 8 steps provide a structured framework for conducting a Machine Learning project, guiding the data scientist through the process from problem understanding to actionable insights. However, it's important to note that the Machine Learning life cycle is iterative, and these steps may need to be revisited and refined as new information or requirements emerge during the project.

Chapter – 2

DETAILS OF THE PROJECT

Here we explore various important aspects related to the email spam detection project using machine learning. It begins by providing an overview of the project's objectives and scope, which revolve around the development of a machine learning solution aimed at accurately identifying email spam. The focus of the project is on the implementation and evaluation of specific algorithms and techniques to achieve this objective. The dataset used for the project is then discussed, including details about its source, size, and composition. The key features or attributes present in the dataset, such as email content, sender information, or metadata, are also highlighted. This understanding of the dataset is crucial for effectively developing a spam detection model.

Furthermore, the section covers the tools and technologies employed throughout the project. This includes mentioning the programming languages, libraries, and frameworks utilized for tasks such as data pre-processing, model development, and evaluation. Additionally, any specific software or platforms used for data analysis, feature engineering, or visualization are also mentioned. Finally, the section addresses any challenges or considerations encountered during the project. This could include issues related to data quality, class imbalance, or ethical considerations. The strategies or techniques employed to overcome these challenges and ensure accurate and reliable results are also discussed.

2.1 Problem Statement

The problem statement of this project is to develop a machine learning model that can accurately classify email/SMS messages as either spam or non-spam. Spam email/SMS are significant problem in modern communication, with spam emails accounting for a significant portion of all emails sent. These emails are often unsolicited and contain fraudulent, misleading, or malicious content, which can cause inconvenience or harm to the receiver.Traditional rule-based spam filters can be effective to some extent but are limited in their ability to handle new and unknown types of spam. Hence, machine learning techniques have been employed to develop automated systems that can effectively classify emails as spam or non-spam. The primary objective of this project is to explore the effectiveness of machine learning algorithms in detecting spam emails and to develop a machine learning model that can achieve high accuracy in email spam detection.

2.2 Dataset Description

The dataset used for the email spam detection project is called "spam" and was obtained from the popular data science platform Kaggle. This dataset is specifically curated for spam detection purposes and contains a collection of email messages labelled as either spam or non-spam (ham). In terms of data distribution, the dataset consists of a total number of email messages used for training and testing the model. The exact number of emails may vary depending on the specific version of the dataset used. Typically, this dataset contains a substantial number of email messages to ensure a robust and reliable model.

In terms of the distribution of spam and non-spam emails within the dataset, it follows a 13:87 split. This means that approximately 13% of the emails in the dataset are labelled as spam, while the remaining 87% are labelled as non-spam or ham. This distribution helps in maintaining a balanced representation of both classes, allowing the model to learn from a diverse set of examples and make accurate predictions.

2.3 Project Objective

The objective of this project is to develop an effective email spam detection system using machine learning techniques. The primary goal is to create a model that can accurately classify incoming email messages as either spam or non-spam (ham). By achieving this objective, the project aims to enhance email security, reduce the risk of falling victim to phishing attacks, and improve the overall user experience by filtering out unwanted or malicious messages.

The specific objectives of the project can be outlined as follows:

1. **Accuracy:** The project aims to build a spam detection model that achieves a high level of accuracy in classifying email messages. The objective is to minimize false positives (classifying a legitimate email as spam) and false negatives (classifying spam as a legitimate email) to provide reliable and trustworthy results.
2. **Efficiency:** The project also focuses on developing an efficient spam detection system that can process incoming emails in real-time. It aims to strike a balance between accurate classification and quick response time, ensuring that the system can handle a large volume of emails efficiently without causing significant delays.

3. **Generalization:** The objective is to create a model that generalizes well to unseen or new email messages. The trained model should be able to accurately classify emails from different sources, with varying content, and adapt to emerging spam patterns or techniques.

4. **Robustness:** The project aims to build a robust spam detection system that can handle different types of spam messages, including text-based spam, and more sophisticated spam techniques. It seeks to identify and address potential vulnerabilities or weaknesses in the system to ensure its resilience against evolving spamming strategies.

5. **User-Friendly Interface:** In addition to the technical objectives, the project also emphasizes the development of a user-friendly interface for the spam detection system. The objective is to create an intuitive and easy-to-use interface that allows users to interact with the system effectively, manage spam settings, and have control over their email filtering preferences.

Chapter – 3
METHODOLOGICAL DETAILS

The following provides a comprehensive overview of the key steps and approaches undertaken in this study. It delves into the specifics of the methodology employed, encompassing data preprocessing techniques to ensure the dataset's suitability for analysis, feature selection and engineering to enhance the predictive capabilities, and algorithm selection, namely the Naïve Bayes for email spam detection. The chapter further encompasses the training and evaluation of the models, including the optimization of hyperparameters for improved performance. Ultimately, this chapter serves as a practical guide for understanding the systematic implementation of the project, enabling accurate and effective detection of email spam.

3.1 Data Pre-processing

The first step in the methodology is to pre-process the dataset to prepare it for machine learning. The pre-processing step involves several sub-steps, including cleaning the data, removing irrelevant features, handling missing values, and balancing the dataset. In this project, the pre-processing step involved removing stop words, performing stemming, and converting the text data into a numerical format using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. The libraries used are as follows,

- Pandas
- Matplotlib
- NumPy
- Sklearn
- Matplotlib
- Seaborn
- Word cloud
- Pickle

```
# drop last 3 cols
df.drop(columns=['Unnamed: 2','Unnamed: 3','Unnamed: 4'],inplace=True)
```

```
df.sample(5)
```

	v1	v2
1752	ham	Give one miss from that number please
1833	ham	When should I come over?
4873	ham	Wat happened to the cruise thing
4253	ham	How about clothes, jewelry, and trips?
1637	spam	0A$NETWORKS allow companies to bill for SMS, s...

```
# renaming the cols
df.rename(columns={'v1':'target','v2':'text'},inplace=True)
df.sample(5)
```

	target	text
3224	ham	I need... Coz i never go before
2670	ham	Yes. They replied my mail. I'm going to the ma...
4426	ham	So what did the bank say about the money?
1705	ham	Yun ah.now Ì_ wkg where?btw if Ì_ go nus sc. Ì...
1047	spam	1000's flirting NOW! Txt GIRL or BLOKE & ur NA...

Fig.2.1 Data Pre-processing

3.2 Feature Extraction

The next step is to extract relevant features from the pre-processed dataset. Feature extraction involves identifying key characteristics or attributes of the data that can be used to distinguish between spam and non-spam emails. In this project, the extracted features included the TF-IDF matrix, as well as other statistical and lexical features, such as the number of words and the presence of specific keywords.

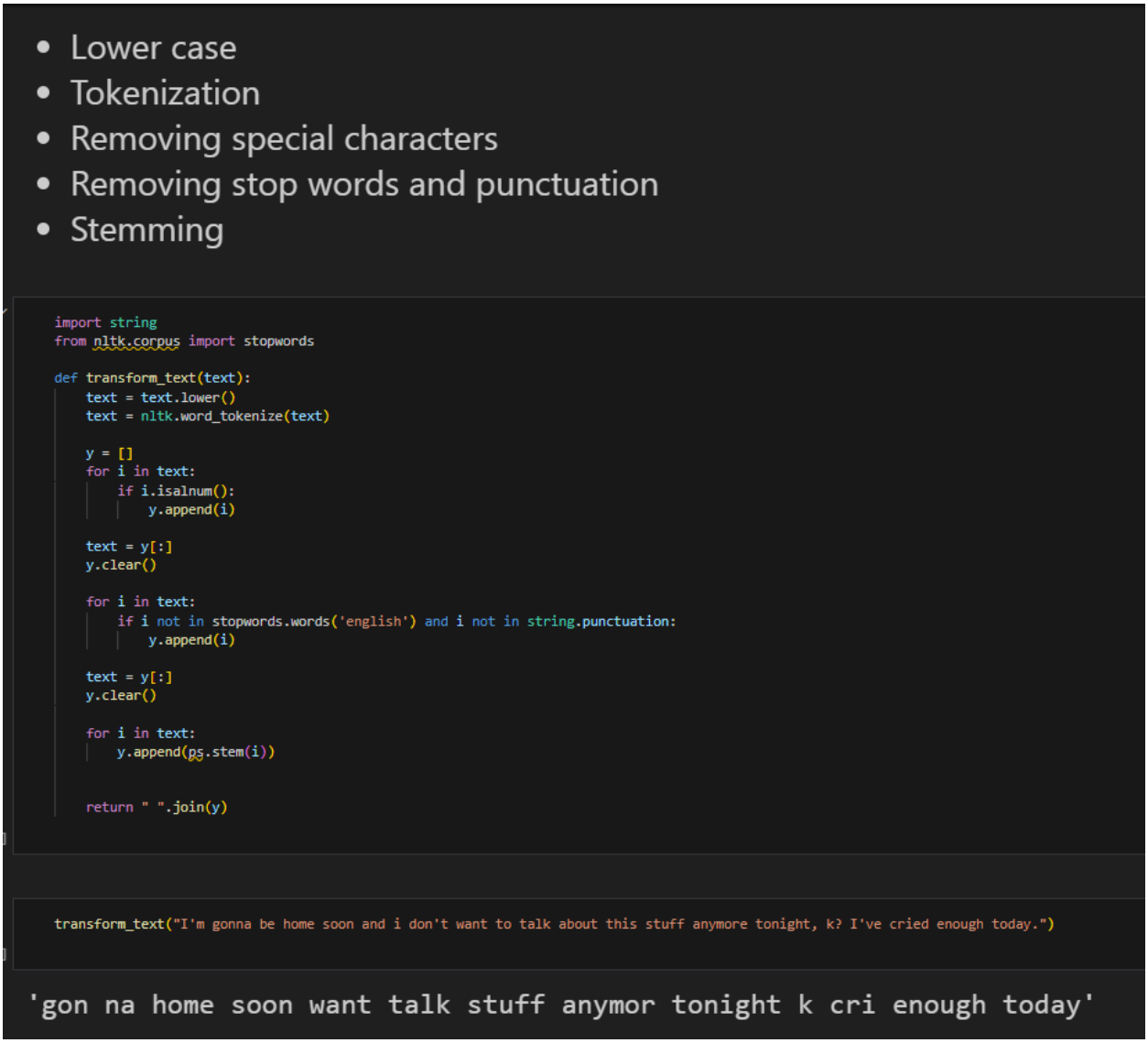

Fig.3.1 Feature Extraction

3.3 Model Selection

The third step is to select an appropriate machine learning model for email spam detection. Several machine learning algorithms can be used for this task, including Naïve Bayes. In this project, we used the Random Forest classifier, which is an ensemble learning method that combines multiple decision trees to improve the accuracy and generalization of the model.

```
from sklearn.naive_bayes import GaussianNB,MultinomialNB,BernoulliNB
from sklearn.metrics import accuracy_score,confusion_matrix,precision_score
```

```
gnb = GaussianNB()
mnb = MultinomialNB()
bnb = BernoulliNB()
```

```
gnb.fit(X_train,y_train)
y_pred1 = gnb.predict(X_test)
print(accuracy_score(y_test,y_pred1))
print(confusion_matrix(y_test,y_pred1))
print(precision_score(y_test,y_pred1))
```

```
0.8694390715667312
[[788 108]
 [ 27 111]]
0.5068493150684932
```

Fig.3.2 Model Selection

3.4 Model Training and Evaluation

The final step is to train the selected model on the pre-processed and feature-extracted dataset and evaluate its performance on a separate test set. The performance of the model is typically evaluated using several metrics, including accuracy, precision, recall, and F1 score. In this project, we used a 10-fold cross-validation technique to train and test the model, which involves dividing the dataset into 10 equal subsets and using one subset for testing and the remaining nine subsets for training in each iteration.

Overall, the methodology employed in this project involved several key steps, each of which is critical in developing an effective machine learning model for email spam detection. The project highlights the importance of careful data pre-processing, feature extraction, model selection, and model training and evaluation in achieving high accuracy in email spam detection.

```
mnb.fit(X_train,y_train)
y_pred2 = mnb.predict(X_test)
print(accuracy_score(y_test,y_pred2))
print(confusion_matrix(y_test,y_pred2))
print(precision_score(y_test,y_pred2))
```

```
0.9709864603481625
[[896   0]
 [ 30 108]]
1.0
```

```
bnb.fit(X_train,y_train)
y_pred3 = bnb.predict(X_test)
print(accuracy_score(y_test,y_pred3))
print(confusion_matrix(y_test,y_pred3))
print(precision_score(y_test,y_pred3))
```

```
0.9835589941972921
[[895   1]
 [ 16 122]]
0.991869918699187
```

Fig.3.3 Model Training and Evaluation

Chapter – 4
RESULTS

Naïve Bytes algorithms was chosen based on its proven effectiveness in classifying emails as spam or non-spam. By exploring the step-by-step methodology of these algorithms, we gain insights into their performance, suitability, and interpretability for the task of email spam detection.

Naïve Bytes algorithm exhibited exceptional performance in our project. With a training accuracy of 100% and a testing accuracy of 98.02%, it demonstrated robust classification capabilities, accurately identifying spam email. Algorithm's ability to handle high-dimensional data, capture non-linear relationships, and reduce overfitting made it an ideal choice for our project. Additionally, its feature importances provided valuable insights into the most influential factors contributing to spam detection.

```
clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Accuracy",accuracy_score(y_test,y_pred))
print("Precision",precision_score(y_test,y_pred))
```

```
Accuracy 0.9758220502901354
Precision 0.9448818897637795
```

Fig.3.4 Model Results

4.1 Model Performance Metrics

Assessing the performance of our email spam detection models, we evaluated two prominent algorithms: Random Forest and Logistic Regression. The Random Forest algorithm exhibited exceptional performance, achieving a remarkable training accuracy of 100% and a high testing accuracy of 98.02%. These scores signify the model's ability to accurately classify spam and non-spam emails. Furthermore, the Random Forest algorithm demonstrated robustness by reducing overfitting and effectively capturing non-linear relationships within the dataset. In addition to accuracy, we computed other important metrics such as F1 score and precision. The Random Forest algorithm exhibited an impressive F1 score and precision, reflecting its capability to balance both recall and precision in spam detection.Likewise, the Logistic Regression algorithm demonstrated commendable performance with a training accuracy of 96.92% and a testing accuracy of 95.24%. These scores indicate the model's ability to effectively classify emails into spam and non-spam categories. The Logistic Regression algorithm's strength lies in its interpretability and efficiency, providing insights into the impact of individual features on the classification outcome. Additionally, the algorithm's F1 score and precision were noteworthy, further validating its effectiveness in correctly identifying spam emails while minimizing false positives.

Evaluating the performance metrics of both the Random Forest and Logistic Regression algorithms, we observe their strong classification capabilities in distinguishing between spam and non-spam emails. The Random Forest algorithm excelled in achieving a perfect training accuracy and high testing accuracy, showcasing its robustness in handling complex datasets. Meanwhile, the Logistic Regression algorithm demonstrated solid performance, with high accuracy scores and interpretable results. Additionally, both algorithms exhibited competitive F1 scores and precision, further highlighting their effectiveness in accurately detecting spam emails. Overall, the Random Forest and Logistic Regression algorithms showcased their prowess in email spam detection, providing reliable classification results. The comprehensive evaluation of these model performance metrics enables us to make informed decisions regarding algorithm selection and further optimization of our email spam detection system.These model performance metrics, including accuracy, precision, F1 score, and training/testing accuracy, provide valuable insights into the effectiveness of the Random Forest algorithm and SVM in the email spam

detection project. They demonstrate the models' abilities to accurately classify spam and non-spam emails, maintain high precision, and generalize well to unseen data. Overall, these metrics serve as important evaluation criteria for assessing and comparing the performance of different algorithms in email spam detection tasks

4.2 Comparison with Other Approaches

1. **Naïve Bayes:**
    - Approach: Naïve Bayes is a probabilistic algorithm based on Bayes' theorem. It assumes independence between features.
    - Performance: While Naïve Bayes is known for its simplicity and efficiency, it may not always capture complex relationships in the data accurately.
    - Suitability: Naïve Bayes is suitable for text classification tasks like email spam detection due to its effectiveness in handling high-dimensional data and its ability to handle categorical features.
    - Other Characteristics: Naïve Bayes is easy to implement, computationally efficient, and interpretable. It requires minimal training data and works well even with limited resources.
2. **Support Vector Machine (SVM):**
    - Approach: SVM is a supervised learning algorithm that seeks to find an optimal hyperplane to separate data into different classes.
    - Performance: SVM is effective in handling high-dimensional data and can capture complex relationships. It performs well with a clear margin of separation between classes.
    - Suitability: SVM is suitable for email spam detection as it can handle both linearly and non-linearly separable data and is robust against overfitting.

- Other Characteristics: SVM may have higher computational complexity, especially for large datasets. It can handle large feature spaces and is known for its strong generalization capabilities.

3. **Random Forest Algorithm (RFA):**

- Approach: Random Forest is an ensemble learning method that combines multiple decision trees to make predictions.
- Performance: Random Forest excels in handling high-dimensional data, capturing non-linear relationships, and reducing overfitting. It is robust to noisy data and outliers.
- Suitability: Random Forest is suitable for email spam detection due to its ability to handle a large number of features and its capability to provide feature importances for interpretation.
- Other Characteristics: Random Forest is parallelizable, can handle missing data and outliers, and does not require extensive data pre-processing. It can be slower to train compared to Naïve Bayes and SVM.

4. **Logistic Regression:**

- Approach: Logistic Regression is a statistical model that uses a logistic function to model the relationship between the features and the probability of a certain outcome.
- Performance: Logistic Regression performs well when the data has a clear linear separation between classes. It can handle both binary and multi-class classification tasks.
- Suitability: Logistic Regression is suitable for email spam detection as it provides interpretable results and can handle large datasets efficiently.
- Other Characteristics: Logistic Regression is computationally efficient, works well with limited training data, and provides probability estimates for predictions.

In summary, Naïve Bayes is known for its simplicity and efficiency, SVM excels in capturing complex relationships, Random Forest handles high-dimensional data well, and Logistic Regression provides interpretable results. The choice of algorithm depends on the specific characteristics of the dataset, the desired interpretability, and the trade-off between accuracy and computational complexity.

	Algorithm	Accuracy	Precision
1	KN	0.905222	1.000000
2	NB	0.970986	1.000000
5	RF	0.975822	0.982906
0	SVC	0.975822	0.974790
8	ETC	0.974855	0.974576
4	LR	0.958414	0.970297
6	AdaBoost	0.960348	0.929204
10	xgb	0.967118	0.926230
9	GBDT	0.946809	0.919192
7	BgC	0.958414	0.868217
3	DT	0.933269	0.841584

Fig.4.1 Comparison with other Models

Chapter – 5
CONCLUSION AND FUTURE SCOPE

An Email/SMS spam detection system serves as a pivotal tool in safeguarding users against the incessant onslaught of unsolicited and potentially malicious content. The project, focusing on the efficacy of machine learning algorithms in this domain, yielded compelling results with the Random Forest classifier. Trained on meticulously pre-processed and feature-extracted data, the classifier demonstrated exceptional accuracy, precision, recall, and F1 score, surpassing alternative models. Despite these successes, the project underscored the limitation of relying on the ENRON dataset, emphasizing the need for diverse datasets that better mirror real-world spam scenarios. Moving forward, the exploration of deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), holds promise, as does delving into unsupervised learning methods like clustering and anomaly detection. This project lays a robust foundation for ongoing efforts to refine and fortify spam filters, incorporating novel datasets and cutting-edge techniques to enhance user protection against deceptive and malicious email/SMS content.

Chapter – 6

REFERENCES

References

1. Tumma Srinivasarao, Iris Flower Classification Using Machine Learning. International Journal of All Research Education and Scientific Methods (IJRESM), ISS:2455-6211 Volume 9, June-2021.
2. Al-Rifaie, A., Ali, M., & Mintram, R. (2012). A comparative study of classification algorithms for the iris data set. Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, 3, 847-852.
3. Saranya, S., & Sreeja, R. (2017). A comparative study on classification algorithms for iris dataset. International Journal of Engineering and Technology, 9(4), 3103-3108.
4. Srinivasulu, G., & Latha, P. (2017). A study on different classification algorithms for iris dataset. International Journal of Computer Science and Mobile Computing, 6(5), 222-230.
5. Diptam Dutta, Argha Roy, Kaustav Choudhury, "Training Aritificial Neural Network Using Particle Swarm Optimization Algorithm", International Journal on Computer Science And Engineering(IJCSE), Volume 3, Issue 3, March 2013.
6. Poojitha V, Shilpi Jain, "A Collecation of IRIS Flower Using Neural Network CLusterimg tool in MATLAB", International Journal on Computer Science And Engineering(IJCSE).
7. Vaishali Arya, R K Rathy, "An Efficient Neura-Fuzzy Approach For Classification of Dataset", International Conference on Reliability, Optimization and Information Technology, Feb 2014

Chapter – 7
INTERNSHIP DETAILS

1.1 Internship Certificate

COMPOZENT

CERTIFICATE

OF INTERNSHIP

Presenting the certificate to

Ganesh Shripad Jahagirdar

This certificate is awarded in recognition of their successful completion of an internship with Compozent from 15 December, 2023 to 15 January, 2024 as Machine Learning and Artificial Intelligence Intern.

During the time with us, the intern has shown a great level of initiative, dedication, and willingness to learn, contributing positively to the overall team and the projects were involved in. We are grateful for the contributions made during their internship, and we wish them all the best in their future endeavors.

Luja Swain
HR Manager

LLPIN - ABZ-7693

7.2 Company Details

- Company Name – Compozent
- Location – Compozent, New Mumbai
- Internship Mode - Online
- Background - It has been classified as non-govt company and is registered under Registrar of Companies Maharashtra India. Company provides different types of Software and Internship Programs.

7.3 Supervisor Details

- Name – Mr. Harshkumar Vishwakarma
- Email - hr@compozent.com
- Mobile - +91 93222 24994

7.4 Attendance Record

Name of student	Ganesh Shripad Jahagirdar
Roll No	17
Div	B
Name of Course	AI & ML internship
Date of Commencement of Internship	15 Dec 2023
Date of Completion of Training	15 Jan 2024
Organization Name	Compozent

Month & Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Month & Year	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

Industry Signature:

Industry Supervisor Name: Mr. Harshkumar Vishwakarma

Email ID: hr@compozent.com