**CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY**


**July to SEPTEMBER 2023**

# Internship Progress Report

MUHAMMAD SAEED ANWAR - BCS203227

ORGANIZATION:

**Eziline Software House Private Limited**

# Table of Contents

**Introduction**.....5

**Objectives…**.....5

**Week 1**.....6

**Day 1:** .....6

**Day 2:** .....6

**Day 3:** .....6

**Day 4:** .....6

**Day 5:** .....6

**Week 2**.....7

**Day 1**: .....7

**Day 2**: .....7

**Day 3**: .....7

**Day 4**: .....7

**Day 5:** .....7

**Week 3**.....8

**Day 1:** .....8

**Day 2**: .....8

**Day 3**: .....8

**Day 4**: .....8

**Day 5:** .....8

**Week 4**.....9

**Day 1:** .....9

**Day 2**: .....9

**Day 3**: .....9

**Day 4**: .....9

**Day 5:** .....9

**Screenshots:**……..…….…..10

**Screenshots:**……..11

**Screenshots:**……..12

**Week 5**.....13

**Day 1:** .....13

**Day 2**: .....13

**Day 3**: .....13

**Day 4**: .....13

**Day 5:** .....13

**Week 6**.....14

**Day 1:** .....14

**Day 2**: .....14

**Day 3**: .....14

**Day 4**: .....14

**Day 5:** .....14

**Week 7**.....15

**Day 1**: .....15

**Day 2**: .....15

**Day 3**: .....15

**Day 4**: .....15

**Day 5:** .....15

**Screenshots:**……..16

**Screenshots:**……..17

**Week 8**.....18

**Day 1**: .....18

**Day 2**: .....18

**Day 3**: .....18

**Day 4**: .....18

**Day 5:** .....18

**Overview…**.....20

**Skills Acquired…**.....20

**Conclusion…...…**.....20

**Conclusion…...…**.....21

## Introduction

I have secured an internship at Eziline Software House Pvt. Ltd. Machine Learning, with a focus on Basics of Machine Learning . The internship period will be from 24th July 2023 to 17th September 2023, and I will be working remotely during this time.

## Objectives

The objective of my internship at Eziline Software House is to gain hands-on experience in Machine Learning, with a primary focus on Machine Learning Techniques. Through active participation in real world projects and collaboration with experienced professionals, I aim to enhance my technical skills, problem-solving abilities, and project management expertise. By contributing to the success of the team's endeavors, I seek to grow as a Machine Learner and make valuable contributions to Eziline's innovative software solutions.

**Week 1:**

**24th July 2023 Day 1:**

I successfully interviewed at Eziline Software House and received an offer letter, securing the internship position.

**25th July 2023 Day 2:**

I was given access to the portal and also they provided me credentials and also taught me how to use.

**25th July 2023 Day 3:**

They assigned me task to get know how of machine learning from internet and I researched online materials regarding to the task. I was given four weeks to complete the project.

**27th July 2023 Day 4:**

I was assigned the Machine learning Task for Spam Email detection.

**28th July 2023 Day 5:**

I read the task thoroughly and started to understand the nature of the task that was assigned to me.

**Week 2:**

**31st July 2023 Day 1:**

I searched different platforms to understand the task. I searched on Youtube, Google, and other Knowledgeful resources.

**01st August 2023 Day 2:**

I started collecting data from different streams regarding to my task. Like Google and Kaggle.

**02nd August 2023 Day 3:**

I started to preprocess the data that I find on the internet and then start applying Techniques.

**03rd August 2023 Day 4:**

After the preprocessing step on day 4 I applied the feature selection algorithm, in my scenario that was best first feature selection algorithm.

**04th August 2023 Day 5:**

On the last day of 2nd week I started studying spam filter Algorithms.

**Week 3:**

**07th August 2023 Day 1:**

I dedicated my day to work on and understand the algorithm NaiveBayes and then tried to apply that.

**08th August 2023 Day 2:**

I worked on and understand the algorithm NaiveBayesMultinomial and then tried to apply that.

**09th August 2023 Day 3:**

I worked on and understand the algorithm J48 Algorithm and then tried to apply that**.**

**10th August 2023 Day 4:**

I trained the model and then started checking the accuracy of the models.

**11th August 2023 Day 5:**

I tested the data and train the models again then compared the results of the algorithms.

**Week 4:**

**14th August 2023 Day 1:**

Pakistan Independence Day Holiday.

**15th August 2023 Day 2:**

I dedicated the day to preparing and evaluating training and testing datasets, fine-tuning models to achieve superior performance.

**16th August 2023 Day 3:**

I compared the results of our spam email detection models, analyzing their precision, recall, and overall effectiveness to refine our email filtering system.

**17th August 2023 Day 4:**

On this day I did Data splitting and evaluation and then did the comparative analysis on my task.

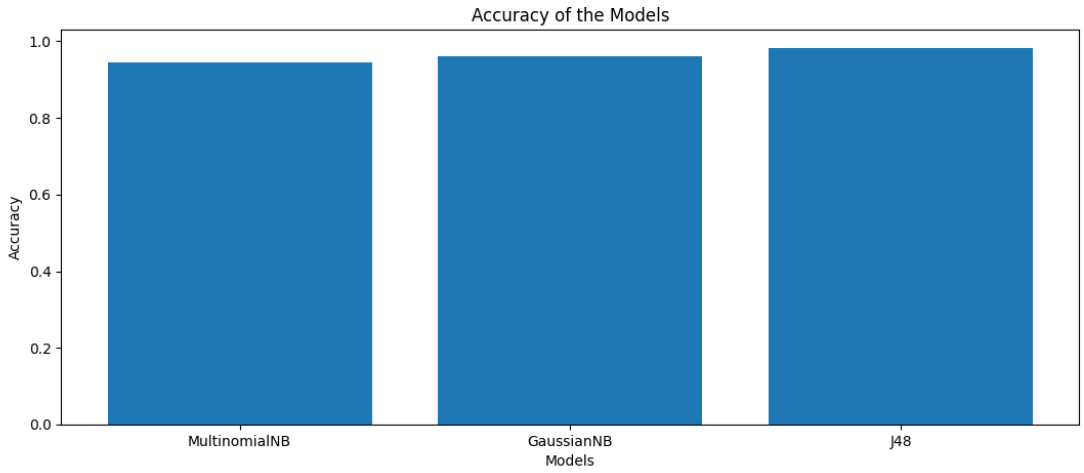
**18th August 2023 Day 5:**

I showed my work to Project Manager.

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, ConfusionMatrixDisplay

df = pd.read_csv('emails.csv')

print(df.describe())

x = df.iloc[:, 1:3001]
y = df.iloc[:, -1].values

train_x, test_x, train_y, test_y = train_test_split(x, y, test_size = 0.3)


from sklearn.naive_bayes import MultinomialNB
mnb = MultinomialNB()
mnb.fit(train_x, train_y)
mnb_y_pred = mnb.predict(x)
mnb_accuracy = accuracy_score(y, mnb_y_pred)
print("Accuracy of MultinomialNB model:",round(mnb_accuracy * 100,2),"%")
```

```
from sklearn.naive_bayes import MultinomialNB
mnb = MultinomialNB()
mnb.fit(train_x, train_y)
mnb_y_pred = mnb.predict(x)
mnb_accuracy = accuracy_score(y, mnb_y_pred)
print("Accuracy of MultinomialNB model:",round(mnb_accuracy * 100,2),"%")
print(confusion_matrix(y, mnb_y_pred))


from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
gnb.fit(train_x, train_y)
gnb_y_pred = gnb.predict(x)
gnb_accuracy = accuracy_score(y, gnb_y_pred)
print("Accuracy of GaussianNB model:", round(gnb_accuracy * 100,2),"%")
print(confusion_matrix(y, gnb_y_pred))


from sklearn import tree
j48 = tree.DecisionTreeClassifier()
j48 = j48.fit(train_x, train_y)
j48_y_pred = j48.predict(x)
j48_accuracy = accuracy_score(y, j48_y_pred)
```

```
from sklearn import tree
j48 = tree.DecisionTreeClassifier()
j48 = j48.fit(train_x, train_y)
j48_y_pred = j48.predict(x)
j48_accuracy = accuracy_score(y, j48_y_pred)
print("Accuracy of J48 model:",round(j48_accuracy * 100,2),"%")
print(confusion_matrix(y, j48_y_pred))

models = pd.DataFrame({
    'Model': ["MultinomialNB", "GaussianNB","J48"],
    'Accuracy':  [mnb_accuracy, gnb_accuracy, j48_accuracy]
})

plt.figure(figsize = (15,5))
plt.bar(models["Model"], models["Accuracy"])
plt.xlabel("Models")
plt.ylabel("Accuracy")
plt.title("Accuracy of the Models")
plt.show()
```



**Week 5:**

**21th August 2023 Day 1:**

I was assigned a new task. It was on the Supervised Machine Learning and Unsupervised Machine Learning. I was given 4 weeks to complete the project.

**22nd August 2023 Day 2:**

On the first day, I started working on the project. I initiated the research on the topics and then I started understanding the topics.

**23rd August 2023 Day 3:**

I work on the supervised Machine Learning. In supervised machine learning, I focused on training predictive models using labeled data. This involved splitting the dataset into training and testing sets, selecting appropriate features.

**24th August 2023 Day 4:**

In regression analysis, I worked on predicting continuous numerical values such as sales figures or temperatures. This involved selecting relevant features, training models like Linear Regression, and assessing performance using metrics like Mean Squared Error (MSE).

**25th August 2023 Day 5:**

For classification tasks, I tackled the challenge of categorizing data into distinct classes, such as spam or not spam emails. I employed algorithms like Random Forest, Naive Bayes in order to achieving high accuracy and precision through model evaluation.

**Week 6:**

**28th August 2023 Day 1:**

In forecasting, I focused on predicting future values based on historical data, particularly in financial or demand forecasting. I utilized methods like ARIMA to make accurate predictions, optimizing models for various forecasting scenarios.

**29th August 2023 Day 2:**

I dedicated my day to worked on unsupervised learning. On the first day, I delved into clustering techniques like K-Means and Hierarchical Clustering. I started with data preprocessing, selecting the appropriate number of clusters using methods like implemented K-Means clustering on a dataset.

**30th August 2023 Day 3:**

Day two was dedicated to association rules mining. I began with data preprocessing, transforming transaction data into the required format. I then applied the Apriori algorithm to uncover frequent item sets and association rules.

**31st August 2023 Day 4:**

This day was centered around dimension reduction using Principal Component Analysis (PCA). I began by understanding the concept of PCA and its applications. I then implemented PCA on a high-dimensional dataset, reducing its dimensionality while preserving as much variance as possible. Visualization techniques were used to illustrate the impact of dimension reduction.

**1st September 2023 Day 5:**

I practiced these techniques on different datasets, comparing their strengths and weaknesses. The day also involved evaluating the clustering results using silhouette scores and visualizations.

**Week 7:**

**4th September 2023 Day 1:**

On the 1st day of week 7, I applied the knowledge gained in association rules mining and dimension reduction to real-world datasets.

**5th September 2023 Day 2:**

I started getting different data sets and then I applied the supervised Machine Learning technique. Firstly, the Regression.

**6th September 2023 Day 3:**

I worked on the model by applying the classification technique on the problem.

**7th September 2023 Day 4:**

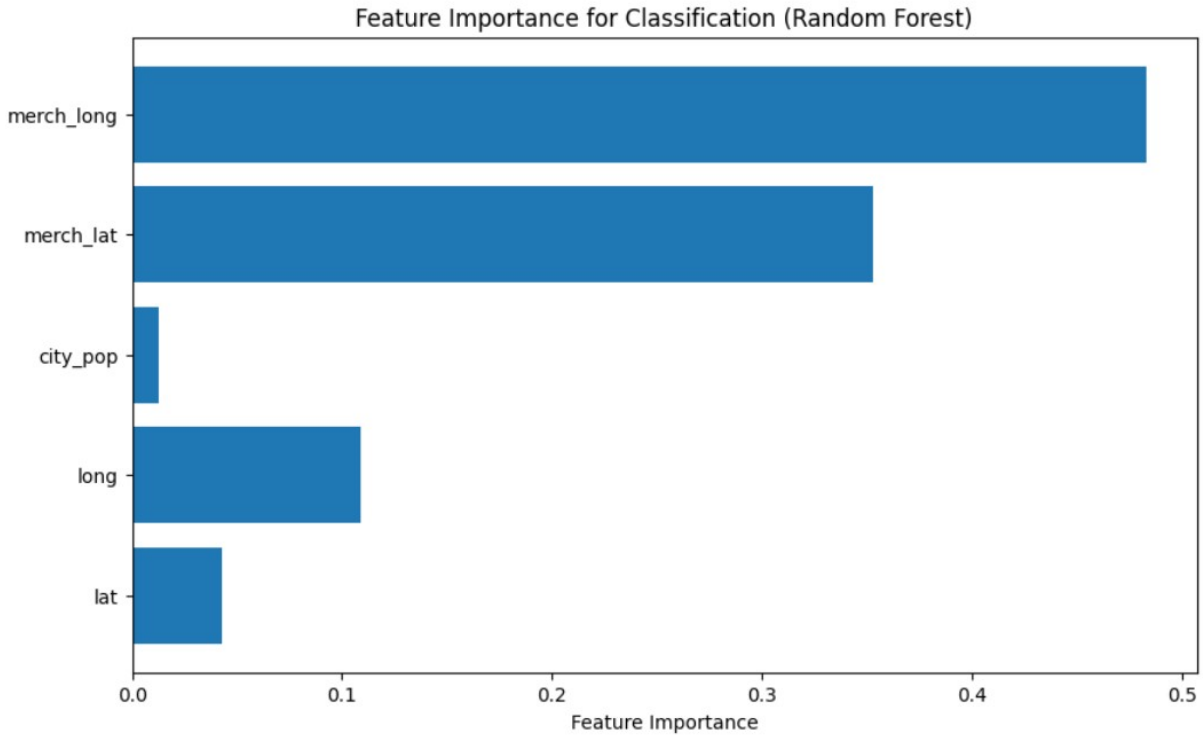
I worked on the model by applying the forecasting technique on the problem and then Analyzed the model accuracy.

**8th September 2023 Day 5:**

Presented my work to the Project Manager. Furthermore, my project manager asked me to enhance my proficiency in data preprocessing. I gave next 2 weekend days on that.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from statsmodels.tsa.arima.model import ARIMA  # Corrected import
import joblib

# Step 1: Load the dataset
data = pd.read_csv('datacamp_workspace_export.csv')

# Assuming the target variable for classification is 'is_fraud'
data = data.dropna(subset=['is_fraud'])

# Step 3: Split data into features (X) and target variables (y_class, y_amount)
# Exclude non-numeric columns for both classification and regression
numeric_columns = ['lat', 'long', 'city_pop', 'merch_lat', 'merch_long']
X_cls = data[numeric_columns]
X_reg = data[numeric_columns]
y_class = data['is_fraud']
```

```
Classification Accuracy: 0.9940029985007496
Confusion Matrix:
[[662   1]
 [  3   1]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       663
           1       0.50      0.25      0.33         4

    accuracy                           0.99       667
   macro avg       0.75      0.62      0.67       667
weighted avg       0.99      0.99      0.99       667

Regression Mean Squared Error: 11416.346164647262
Regression R-squared: 0.02380273497758889
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: Fut
  warnings.warn(
['credit_card_fraud_regression.pkl']
```


Feature Importance for Classification (Random Forest)
merch_long
merch_lat
city_pop
long
lat
0.0
0.1
0.2
0.3
0.4
0.5
Feature Importance

**Week 8:**

**11th September 2023 Day 1:**

Presented my work to the project manager again.

**12th September 2023 Day 2:**

I started studying the unsupervised machine learning.

**13th September 2023 Day 3:**

Tried to apply dataset for unsupervised but it was only for supervised machine learning.

**14th September 2023 Day 4:**

I started preparing my Report.

**15th September 2023 Day 5:**

Received an Internship Certificate from the Project Manager at Eziline Software House.

### Overview:

During my internship, I actively engaged in projects related to spam email detection, supervised machine learning, and unsupervised machine learning. These projects provided me with hands-on experience in the field of machine learning, allowing me to apply theoretical knowledge to real-world scenarios. I was responsible for tasks such as sourcing and preprocessing datasets, training and fine-tuning machine learning models, and conducting comprehensive comparisons among various algorithms. These experiences enhanced my understanding of machine learning techniques, sharpened my data preprocessing skills, and provided valuable insights into the nuances of model evaluation and selection. Overall, my internship was a rich learning journey that significantly advanced my expertise in machine learning.

### Skills Acquired:

During my internship as a Machine Learning intern at Eziline Software House, I had the privilege to acquire and strengthen a versatile skill set crucial in the domain of machine learning. These hands-on experiences not only facilitated the growth of my technical prowess but also granted me valuable insights and effective problem-solving.

## Conclusion:

### Experience at Eziline Software House:

During my internship at Eziline Software House, I had a highly positive experience, particularly in the field of Machine Learning. I became adept at various Machine Learning techniques and gained hands-on experience in model training, data preprocessing, and result comparisons. While I didn't encounter major challenges, the guidance from my project manager was invaluable, helping me gain a deeper understanding of machine learning projects.

The constructive feedback I received further motivated me to improve and refine my work in Machine Learning. I am now enthusiastic about continuing my journey in this field, embracing new challenges, and expanding my skill set in the realm of Machine Learning. This internship has been immensely insightful, and I look forward to further exploration and growth in the fascinating world of machine learning.