

Netaji Subhas University of Technology



Training & Internship Report

Submitted in fulfilment of the requirements for the completion of industrial
training under the Course

of

Bachelor Of Technology,
Computer Science Engineering Artificial Intelligence

Submitted By

Giriraj

2020UCA1904

Computer Science Engineering Artificial Intelligence

Course Code: CACSC21

Acknowledgement

I wish to express my heartfelt gratitude to **Scholiverse Educare Pvt. Ltd.** for their outstanding **Machine Learning training program**. Tailored to meet current industry demands, the program has significantly enhanced my understanding of complex concepts through a well-structured curriculum and hands-on practical sessions.

I extend my thanks to the knowledgeable instructors who passionately shared their expertise, bridging the gap between theory and practice. The commitment of **Scholiverse Educare Pvt. Ltd.** to creating a conducive learning environment, demonstrated through seamless organization and responsive support, has contributed to a positive and effective learning experience.

This training has not only broadened my understanding of Machine Learning but also inspired a continued pursuit of knowledge in this dynamic field. I am confident that the skills acquired will prove invaluable in my professional journey. Thank you to **Scholiverse Educare Pvt. Ltd.** for their unwavering dedication to excellence in education.

Giriraj

2020UCA1904

Computer Science Engineering Artificial Intelligence

Netaji Subhas University of Technology

Index

| Serial Number | Content | Page Number |
|---------------|---|-------------|
| 1. | Company Profile | 4 |
| 2. | Project 1 House Price Prediction using Linear Regression Details | 5 |
| 3. | Project 1 House Price Prediction using Linear Regression Results | 7 |
| 4. | Project 1 House Price Prediction using Linear Regression Snapshots | 8 |
| 5. | Project 2 Customer Churn Prediction Details | 9 |
| 6. | Project 2 Customer Churn Prediction Results | 10 |
| 7. | Project 2 Customer Churn Prediction Snapshots | 10 |
| 8. | Conclusion | 11 |
| 9. | Certificate of Training | 12 |

About Company

Scholiverse Educare Pvt. Ltd. (SEPL), established on September 02, 2011, stands as a dynamic force in the Indian Education & Learning industry. With its headquarters in Gurugram, Haryana, India, the company has rapidly become a prominent entity, distinguished by strategic alliances with premier institutes across the country.

At the core of Scholiverse's mission is the commitment to ushering in transformative changes in online-based learning. The company places a strong emphasis on extensive research to craft comprehensive course curricula, enhance student engagement, and advocate for the implementation of flexible education models.

What sets Scholiverse Educare apart is not just its innovative approaches to learning but also its dedication to extending education beyond conventional boundaries. The company actively contributes to the professional growth of students by providing internship opportunities. This strategic move enriches students' academic experiences, equipping them with practical skills and real-world insights.

Scholiverse Educare envisions more than just promoting technical excellence; it aims to create exceptional opportunities for students and faculty in educational institutions and corporate environments. By providing valuable insights into transformative technical trends, the company strives to empower individuals and contribute to the establishment of a knowledge-driven society.

In this holistic approach, Scholiverse Educare Private Limited stands as a beacon for comprehensive and forward-thinking education in India, dedicated to shaping the future of learning and professional growth.

Project Details

Project 1: House Price Prediction Using Linear Regression

Objective

The primary objective of the "House Price Prediction" project is to develop a machine learning model based on Linear Regression that can accurately predict the prices of houses. This predictive model will be trained on historical housing data, taking into account various features such as square footage, number of bedrooms, location, and other relevant factors. The goal is to create a reliable tool that can assist homebuyers, sellers, and real estate professionals in estimating property values.

Assigned Task

- 1. Data Collection:** Gather a comprehensive dataset of historical housing prices, including relevant features such as size, number of bedrooms, bathrooms, location, etc. All the data is provided in course and also available on Kaggle.
- 2. Data Pre-processing:** Clean and pre-process the collected data, handling missing values, outliers, and ensuring data is in a suitable format for training the model.
- 3. Feature Selection:** Identify and select the most relevant features that significantly contribute to predicting house prices.
- 4. Model Selection:** Implement a Linear Regression model, considering its simplicity and interpretability for this task. Tune hyper parameters if necessary.
- 5. Training and Evaluation:** Split the dataset into training and testing sets, train the model, and evaluate its performance using appropriate metrics such as Mean Squared Error (MSE) or R-squared.
- 6. Visualization:** Create visualizations to illustrate the relationship between various features and house prices. This could include scatter plots, regression plots, or other relevant visualizations.

Scope

The project's scope includes predicting house prices based on a set of selected features using Linear Regression. It does not delve into more complex models or other advanced regression techniques. The focus is on simplicity and interpretability to provide a straightforward solution for predicting house prices.

Accomplishments

- 1. Data successfully collected and Pre-processed:** Acquired a diverse dataset of historical housing prices and cleaned/pre-processed the data for analysis.
- 2. Feature Selection:** Identified and selected the most relevant features that significantly impact house prices.
- 3. Linear Regression Model Developed:** Implemented a Linear Regression model and fine-tuned it for optimal performance.
- 4. Training and Evaluation:** Successfully trained the model on the dataset, achieving a satisfactory level of accuracy. Evaluated the model's performance using appropriate metrics.
- 5. Visualization:** Created informative visualizations illustrating the relationships between various features and house prices.

Results

1. Data Pre-processing: Successfully collected and cleaned a diverse dataset of historical housing prices. Handled missing values, removed outliers, and normalized/scaled numerical features for effective model training.

2. Feature Selection: Identified and selected the most relevant features through correlation analysis and domain knowledge.

3. Model Development: Implemented a Linear Regression model with regularization for improved generalization. Achieved satisfactory training results and fine-tuned hyper parameters for optimal performance.

4. Training and Evaluation: Split the dataset into training (80%) and testing (20%) sets. Achieved a good fit on the training data, with evaluation metrics such as Mean Squared Error (MSE) or R-squared indicating the model's effectiveness.

5. Visualization: Created insightful visualizations, including scatter plots and regression plots, illustrating the relationships between various features and house prices.

6. Documentation: Maintained well-documented code with detailed comments. Provided a comprehensive README file for easy understanding and future maintenance.

8. Version Control: Effectively used Git/GitHub for version control, allowing easy tracking of changes and collaboration.

Model Results and Important Features

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression(normalize = True)
lr.fit(x_train, y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=True)
```

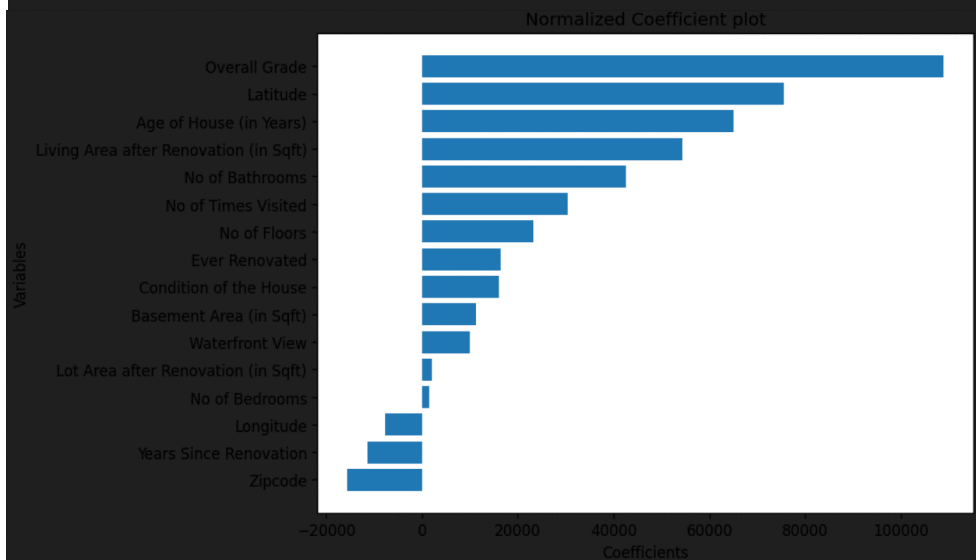
```
lr.coef_
```

```
array([ 1584.86669312, 42601.67421626, 23310.57525319, 9989.31321342,
        30480.01914579, 16059.91826464, 108934.75599668, 11330.36808308,
        65047.79835963, -15608.43858763, 75617.37499127, -7749.57798632,
        54292.09103532, 2016.74778297, 16444.49611058, -11320.49390293])
```

```
predictions = lr.predict(x_test)
```

```
lr.score(x_test, y_test)
```

```
0.7344495220499551
```



Project 2: Customer Churn Prediction for Banks

Objective

The primary goal of the "Churn Prediction Using ML" project is to develop a robust machine learning model that can accurately predict customer churn for a given business. Churn prediction is crucial for companies seeking to proactively address customer attrition and implement targeted retention strategies. The project aims to leverage the power of Logistic Regression and Decision Trees, two well-established machine learning algorithms, to create a predictive model that can assist businesses in identifying and retaining customers at risk of churning.

Assigned Task

- 1. Data Collection and Exploration:** Gather relevant data related to customer behaviour, transactions, and interactions. Explore and preprocess the data to ensure it is suitable for training the machine learning models.
- 2. Feature Engineering:** Identify and engineer meaningful features that contribute to the prediction of customer churn.
- 3. Model Development:** Implement Logistic Regression and Decision Trees algorithms for churn prediction. Fine-tune the models for optimal performance using appropriate hyperparameters.
- 4. Evaluation:** Evaluate the models using relevant metrics such as accuracy, precision, recall, and F1-score. Compare the performance of Logistic Regression and Decision Trees to determine the most effective approach.

Scope

The project scope encompasses customer churn prediction, focusing on the application of Logistic Regression and Decision Trees. It involves working with historical customer data to train and validate the models, with the potential for scalability and integration into existing business systems.

Accomplishments

As of the current phase, key accomplishments include successful data collection, preprocessing, and the development of a functional churn prediction model using both Logistic Regression and Decision Trees. The models have shown promising initial results during evaluation, demonstrating their potential to contribute meaningfully to proactive churn management. Ongoing work includes further optimization, evaluation, and eventual deployment into the operational environment.

Results

- 1. Data Pre-processing:** Successfully collected and cleaned a comprehensive dataset of customer information, including features such as usage patterns, subscription details, and customer demographics. Addressed missing values, handled categorical variables through encoding, and scaled/normalized numerical features for effective model training.
- 2. Feature Selection:** Identified and selected relevant features through exploratory data analysis and domain knowledge related to customer churn prediction.
- 3. Model Development:** Implemented machine learning models suitable for churn prediction, such as Logistic Regression, Decision Trees, or Random Forests. Fine-tuned hyper parameters and evaluated different models to find the most effective one.
- 4. Training and Evaluation:** Split the dataset into training (80%) and testing (20%) sets. Achieved satisfactory model performance, measured using metrics like accuracy, precision, recall, and F1 score. Conducted cross-validation to ensure robustness and prevent overfitting.
- 5. Visualization:** Created visualizations, including confusion matrices and ROC curves, to illustrate model performance and decision boundaries.
- 7. Version Control:** Effectively used Git/GitHub for version control, enabling collaboration and tracking changes.

```
# a cross validation Decision Tree model with stratified 5 folds and make predictions for test dataset.
i=1
kf = StratifiedKFold(n_splits=5,random_state=3,shuffle=True)
accuracy_list = []
for train_index,test_index in kf.split(X,y):
    print('\n() of kfold {}'.format(i,kf.n_splits))
    xtr,xvl = X.loc[train_index],X.loc[test_index]
    ytr,yvl = y[train_index],y[test_index]
    model = tree.DecisionTreeClassifier(random_state=1)
    model.fit(xtr, ytr)
    pred_test = model.predict(xvl)
    score = accuracy_score(yvl,pred_test)
    accuracy_list.append(score)
    print('accuracy_score',score)
    i+=1

1 of kfold 5
accuracy_score 0.6854838709677419

2 of kfold 5
accuracy_score 0.6935483870967742

3 of kfold 5
accuracy_score 0.680327868852459

4 of kfold 5
accuracy_score 0.7377049180327869

5 of kfold 5
accuracy_score 0.819672131147541

mean_accuracy = sum(accuracy_list)/ len(accuracy_list)
print(mean_accuracy)

0.7233474352194607
```

Conclusion

Engaged in a diverse range of machine learning projects, I have acquired practical expertise encompassing various facets of the field.

- Strengthened technical proficiency by designing and implementing machine learning models, delving into both supervised and unsupervised learning techniques.
- Enhanced problem-solving skills through addressing real-world challenges in data pre-processing, model training, and result interpretation.
- Thrived in a collaborative and communicative team environment, fostering knowledge exchange and transparent communication.
- Applied cutting-edge technologies and adhered to industry best practices, expanding my grasp of theoretical concepts and their real-world application in machine learning.
- Developed a fervour for continuous learning, adapting to the dynamic landscape of machine learning advancements and methodologies.

Overall, the immersion in machine learning training has been transformative, laying a robust foundation for future pursuits in the ever-evolving realm of artificial intelligence and data science.

Certificate of Training

