SUPERIOR UNIVERSITY

Programming for Artificial Intelligence – Lab

Task 1

Name: Minhaj Asghar

Roll no: SU92-BSAIM-F23-108

Section: BSAI-4B

Submitted to: Sir Rasikh Ali

House Price Prediction Model Report

1. Introduction

The objective of this project is to develop a model that predicts house prices based on various features such as location, square footage, number of bedrooms, age of the property, and other relevant attributes. Accurate predictions of house prices are valuable for both real estate professionals and potential buyers.

2. Problem Definition

The goal is to predict the price of a house using available data. The challenge is to build a model that can generalize well from the training data and make accurate predictions for unseen data.

3. Dataset Description

The dataset used for this model includes various features:

- **Features:** 'Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street', 'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'SaleType', 'SaleCondition', etc.
- **Target Variable:** SalePrice.

4. Data Preprocessing

Before training the model, the following steps were taken:

- **Handling Missing Values:** Missing data were either imputed or dropped based on the significance of the feature.

Train Data:

```
impute = KNNImputer()

for i in train_df.select_dtypes(include="number").columns:
    train_df[i] = impute.fit_transform(train_df[[i]])
```

```
for i in train_df.select_dtypes(include="object").columns:
    train_df[i] = train_df[i].fillna(train_df[i].mode()[0])
```

Test Data:

```
for i in test_df.select_dtypes(include="number").columns:
    test_df[i] = impute.fit_transform(test_df[[i]])
```

```
for i in test_df.select_dtypes(include="object").columns:
    test_df[i] = test_df[i].fillna(test_df[i].mode()[0])
```

- **Feature Encoding:** Categorical variables (like Street) were converted into numerical representations using techniques such as label encoding.

Train Data:

```
le = LabelEncoder()

for i in train_df.select_dtypes(include="object").columns:
    train_df[i] = le.fit_transform(train_df[i])
```

Test Data:

```
le = LabelEncoder()
for i in test_df.select_dtypes(include="object").columns:
    test_df[i] = le.fit_transform(test_df[[i]])

test_df = test_df.select_dtypes(include="number").astype("int64")
```

5. Model Selection

A Support Vector Classifier (SVC) was used for this classification task. SVC is effective in handling non-linear relationships and works well for multi-class classification problems, which makes it ideal for predicting house price categories.

6. Model Training

The dataset was split into training and testing sets. The Support Vector Classifier (SVC) model was trained using the training data, and predictions were made on the test data. The model was prepared for evaluation based

on standard classification metrics, though further refinement is required to ensure optimal performance.

```
model = SVC()
model.fit(train_x,train_y)
```

```
SVC()
```

7. Challenges and Next Steps

Due to the mismatch in the shapes of the training and testing sets, the model training was not fully evaluated. Possible next steps include:

- Ensuring Consistent Feature Sets: Check for feature alignment between training and testing data before training.
- Fixing Data Preprocessing Issues: Recheck data preprocessing steps, particularly handling missing data and feature encoding, to make sure both sets have the same features.
- Model Re-training and Evaluation: Once the issue is resolved, re-train the model and evaluate it using common classification metrics like accuracy, precision, recall, and F1 score.

8. Submission

Submission of the predictions is in a csv.

```
submission = pd.DataFrame({
    "Id" : test_df.index,
    "SalePrice" : y_pred
})
```

```
submission = submission.to_csv("HouseSalesPrediction.csv")
print("Submission done successfully")
```

```
Submission done successfully
```

9. Conclusion

The Support Vector Classifier (SVC) was selected for the classification task, but due to a mismatch in the shapes of the training and testing sets, the model could not be fully evaluated. The next steps involve resolving data preprocessing issues and ensuring the consistency of feature sets before training the model again. Once resolved, further evaluation will be

performed to determine the effectiveness of the model in predicting house price categories.