Faculty of Computers and Artificial Intelligence Helwan University

Cover sheet

Machine Learning Project

Team no. :

Name	ID
يوسف الأمجد رأفت	20211066
بدر فيصل عبدالرؤوف عبد الحليم	20210219
احمد ايمن كامل مختار	20210030
مارتينا سليمان سامي	20210704
يوسف محمد مجدي كامل	20211102
يوسف ياسر يوسف محمد	20211117

Project Description Document

Model 1: [Support Vector Machine]

General Information on Classification Dataset:

- **Project Description Document Model 1:** Support Vector Machine for Breast Cancer Diagnosis
- **Dataset Name:** Breast Cancer Dataset
- **Number of Classes:** Classification Task with 2 classes M or B (Predicting Diagnosis)
- **Total Number of Samples:** 570
- **Training Samples:** 456
- **Testing Samples:** 114

Implementation Details:

Feature Extraction Phase:

Number of Features Extracted: All features in the dataset

Feature Names:

- Mean radius
- Mean texture
- Mean perimeter
- Mean area
- Mean smoothness
- Mean compactness
- Mean concavity
- Mean concave points
- Mean symmetry
- Mean fractal dimension
- Radius error
- Texture error
- Perimeter error
- Area error
- Smoothness error
- Compactness error
- Concavity error
- Concave points error
- Symmetry error
- Fractal dimension error
- Worst radius
- Worst texture
- Worst perimeter
- Worst area
- Worst smoothness
- Worst compactness
- Worst concavity
- Worst concave points
- Worst symmetry
- Worst fractal dimension

Dimension of Resulted Features: 29 features

Feature Scaling:

- Standardize input features using StandardScaler

Cross-Validation:

- **Used**: No

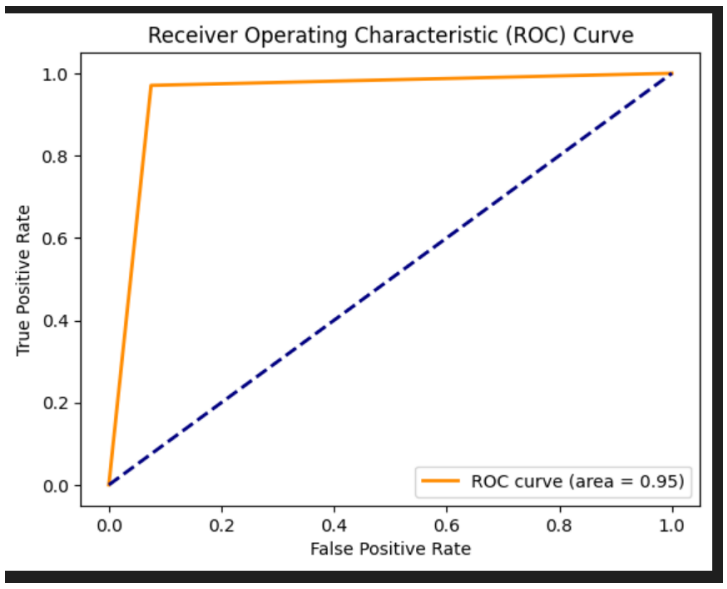
Hyperparameters:

- **Kernel:** Linear
- **Gamma:** Auto
- **Regularization Parameter (C) :** 2

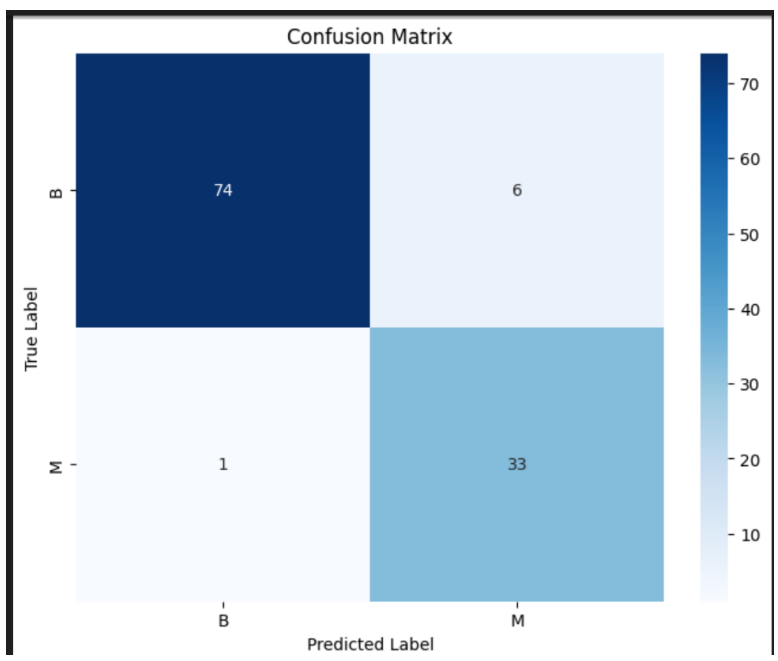
Results Details:

Testing Data:

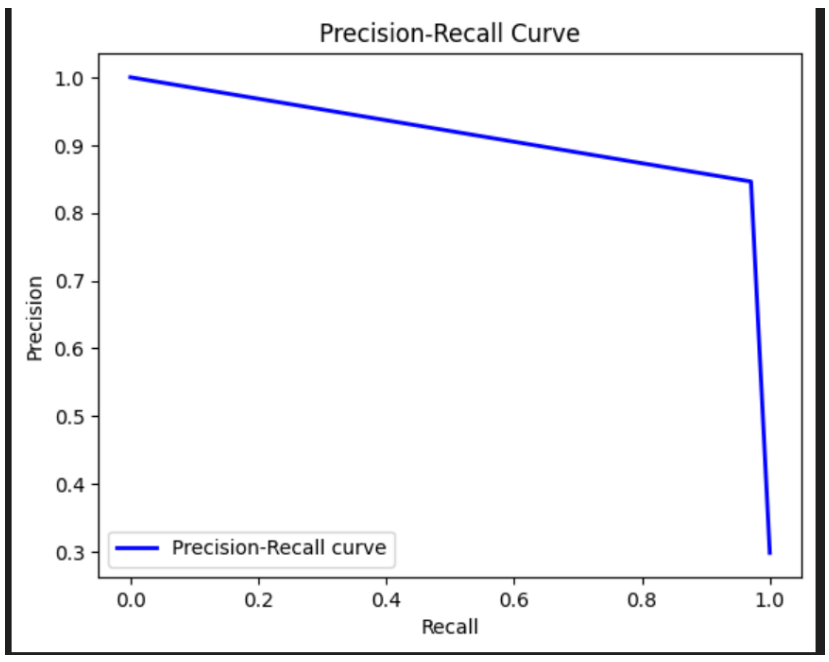
- **ROC Curve:**



- **Confusion Matrix:**



- **Percision Recall Curve:**



- **Accuracy:** [0.93854912280702]
- **ROC AUC Score :** [0.94779411]

Model 2: [Decision Tree]

General Information on Numerical Dataset:

- **Project Description Document Model 2:** Decision Tree Regression for Netflix Stock Price Prediction
- **Dataset Name:** Netflix Stock Price Prediction
- **Number of Classes:** Regression task (Predicting Close Price)
- **Total Number of Samples:** Variable (Depends on the dataset)
- **Training Samples:** Variable (Depends on the dataset)
- **Testing Samples:** Variable (Depends on the dataset

Implementation Details:

Data Preprocessing:

- Handle Missing Values: Drop rows with missing values

Visualization:

- Visualize the Close Price Data

Feature Scaling:

- Used MinMaxScaler to scale the features

Feature Extraction Phase:

Number of Features Extracted: All features in the dataset

Feature Used:

- Open , High , Low , Volume

Target Variable:

- Close Price

Dimension of Resulted Features: 8 features

Cross-Validation:

- **Used**: Yes
- **Number of folds :** 5
- **Training/Validation Ratio/Testing :** 60% / 20% / 20%

Hyperparameter Tuning:

- GridSearchCV to find the best hyperparameters
- Hyperparameters Tuned: max_depth, min_samples_split, min_samples_leaf

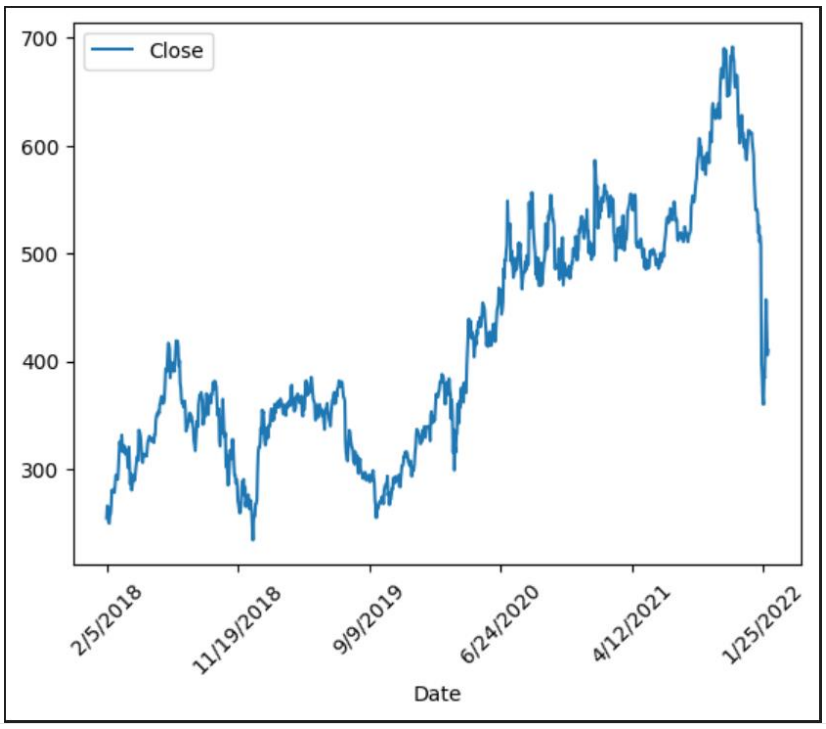
Results Details:

Best Model Parameters:

- max_depth: [Best parameter value]
- min_samples_split: [Best parameter value]
- min_samples_leaf: [Best parameter value]

Evaluation Metrics:

- **Mean Squared Error (MSE) :** [46.2700]
- **Mean Absolute Error (MAE) :** [4.769]
- **Median Absolute Error (MedAE) :** [3.5445]
- **R2 Score :** [0.9960]



Model 3 : [ANN]

General Information on Image Dataset:

- **Project Description Document Model 1:** Artificial Neural Network (ANN) for Netflix Stock Price Prediction
- **Dataset Name:** Netflix Stock Price Dataset
- **Number of Classes:** Regression (Predicting Close Price)
- **Total Number of Samples:** Variable (depends on the dataset)
- **Training Samples:** Variable (depends on the dataset)
- **Testing Samples:** Variable (depends on the dataset)

Implementation Details:

Data Preprocessing:

- Read dataset from 'NFLX.csv'
- Split data into train and test sets using a 80-20 split

Feature Extraction Phase:

Number of Features Extracted: All features in the dataset

Feature Names : Open , High , Low , Volume

- Features Used: Open, High, Low, Volume
- Target Variable: Close Price

Feature Scaling:

- Standardize input features using StandardScaler

Model Architecture:

- Neural Network Architecture:
- Input Layer: 4 neurons (one for each feature)
- Hidden Layers:
- Layer 1: 128 neurons, ReLU activation, 20% dropout
- Layer 2: 64 neurons, ReLU activation, 20% dropout
- Layer 3: 32 neurons, ReLU activation
- Output Layer: 1 neuron (outputting the predicted close price)

Model Compilation:

- Loss Function: Mean Squared Error (MSE)
- Optimizer: Adam with learning rate of 0.001

Model Training:

- Train the model for 100 epochs with a batch size of 32
- Validation split of 20% used for monitoring training progress

Hyperparameters:

- Learning Rate: 0.001
- Dropout Rate: 20%

Results Details:

Training and Validation Loss:

- Plot of Training and Validation Loss over epochs

Predictions:

- Predictions made on the test set

Evaluation Metrics:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Median Absolute Error (MedAE)
- R2 Score

Results:

Mean Squared Error: [47.05521633022602]
Mean Absolute Error: [5.029698599396657]
Median Absolute Error: [4.12031537597656]
R2 score: [0.9958769246324329]

User Interaction:

- Function to predict stock price based on user input for open price, high price, low price, and volume
- User prompted to input open price, high price, low price, and volume
- Predicted close price displayed based on user input