National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Sathvika Bandarapu
Student ID: x23249927

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: …….Sathvika Bandarapu ……

Student ID: ……x23249927……..……

Programme: …Data Analytics…… **Year:** ……2024……..

Module: ……Research Project…….……

Lecturer: …Bharat Agarwal……..……

Submission Due Date: ……12/12/2024……..……

Project Title: ……Enhancing E-commerce Supply Chain and Shipping Efficiency with Machine Learning and Deep Learning Models……

Word Count: ……6993…… **Page Count:** ……19…….……..……

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Sathvika Bandarapu……

Date: ……12/12/2024……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	

Penalty Applied (if applicable):	

Configuration Manual

Sathvika Bandarapu
Student ID: x23249927

1 Project Overview

Despite advancements in machine learning (ML) and deep learning (DL) techniques, current research predominantly focuses on predicting mainly late delivery risk rather than estimating scheduled and actual shipping durations which is very critical aspect for enhancing supply chain resilience. To address the gap, this study aims to predict both scheduled and actual shipping durations using the DataCo SMART SUPPLY CHAIN dataset. This research aims to make predictions as a multi-output regression problem while also employing extensive preprocessing, feature selection, and hyperparameter optimization to mitigate the prediction error.

2 Files

Code is submitted in two files:

1) code_dataco_MLmodels.ipynb: In this file code for ML models are presented
2) code_data__lstmgru.ipynb : In this file code for DL models are presented
3) Dataset with name “DataCoSupplyChainDataset.csv” is also added in the folder.

3 Installations required

Use below command to install all the necessary installations required to run the code:

- **pip install numpy matplotlib tensorflow pandas scikit-learn seaborn scipy xgboost**

4 Steps to be followed

1) Load the dataset

```
# Load the dataset
file_path = 'DataCoSupplyChainDataset.csv'
dataset = pd.read_csv(file_path, encoding='ISO-8859-1')
```

2) Apply label encoder

```
for col in categorical_columns:
    cleaned_dataset[col] = label_encoder.fit_transform(cleaned_dataset[col])
preprocessed_dataset = cleaned_dataset
```

3) Test train validation split

```
# Step 1: Split into training (80%) and testing (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 2: Split 10% of the training data for validation (90% of training for actual training)
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.1, random_state=42)
```

4) Training decision tree model with hyper-parameters

```
# Define the parameter grid for hyperparameter tuning
param_grid = {
    'max_depth': [5, 10, 15, 20, 30],
    'min_samples_split': [2, 5, 10],
    'random_state': [42]
}

# Initialize the model
model = DecisionTreeRegressor()

# Set up GridSearchCV with cross-validation
grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=5, n_jobs=-1, verbose=1, scoring='neg_mean_squared_error')

# Fit the model using the training data (validation happens during grid search)
grid_search.fit(X_train, y_train)
```

5) Similarly train the other ML models

```
# Define the model
rf_model = RandomForestRegressor(random_state=42)

# Define the parameter grid for tuning
param_grid = {
    'n_estimators': [10,20,30,50,100],       # Number of trees in the forest

}

# Initialize GridSearchCV
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=2, n_jobs=-1, verbose=2,scoring='neg_mean_squared_error')

# Fit the GridSearchCV to the training data
grid_search.fit(X_train, y_train)
```

6) Intialise and train the DL models

```
# LSTM model for multi-output regression
lstm_model = Sequential()
# First LSTM layer with return_sequences=True to pass sequences to the next LSTM layer
lstm_model.add(LSTM(50, activation='relu', input_shape=(X_train_scaled.shape[1], X_train_scaled.shape[2]), return_sequences=True))

# Second LSTM layer
lstm_model.add(LSTM(50, activation='relu'))  # Second LSTM layer without return_sequences

# Output layer with 2 neurons (for the 2 target variables)
lstm_model.add(Dense(2))
lstm_model.compile(optimizer='adam', loss='mse')

# Train the model with validation data
history = lstm_model.fit(X_train_scaled, y_train, epochs=10, batch_size=32, validation_data=(X_val_scaled, y_val))

# Predict
y_pred = lstm_model.predict(X_test_scaled)
```

7) All the codes submitted ouput corresponding results and plots

5 System Requirements

- Processor: Minimum or greater than i5 processor required
- RAM of minimum 8GB required
- Code can be executed in visual studio or jupyter notebook or Google colab
- GPU is optional

References

Issaoui, Y., Khiat, A., Bahnasse, A. and Ouajji, H. (2021). An advanced lstm model for optimal scheduling in smart logistic environment: E-commerce case, IEEE Access 9: 126337–126356.

Jana, A. K. (2021). Optimization of e-commerce supply chain through demand prediction for new products using machine learning techniques, J Artif Intell Mach Learn & Data Sci 1(1): 565–569.

Kalkha, H., Khiat, A., Bahnasse, A. and Ouajji, H. (2023). The rising trends of smart e-commerce logistics, IEEE Access 11: 33839–33857.

Mittal, U. and Panchal, D. (2023). Ai-based evaluation system for supply chain vulnerabilities and resilience amidst external shocks: An empirical approach, Reports in Mechanical Engineering 4(1): 276–289.

Sayyad, N. and Varshney, G. (2024). Ai-based hybrid ml models for supply chain optimization, Artificial Intelligence in Supply Chain Management .

Sharma, M., Sharma, V. and Kapoor, R. (2022). Study of e-commerce and impact of machine learning in e-commerce, pp. 1–22.