

Configuration Manual

MSc Research Project
Financial Technology

Abeen Pulikkottil Thambi
Student ID: x21224714

School of Computing
National College of Ireland

Supervisor: Victor del Rosal

National College of Ireland
MSc Project Submission Sheet



School of Computing

Student Name:Abeen Pulikkottil Thambi.....
Student ID:x21224714.....

Programme:Msc FinTech..... **Year:**2022/23.....

Module: ...Research Project.....

Lecturer: ...Victor del Rosal.....

Submission Due Date:13th of August 2023.....

Project Title: Factors Influence Customer Satisfaction of International Remittances / International Money Transfers Services Using Ensemble Machine Learning.....

Word Count:1579..... **Page Count:**10.....

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:ABEEN PULIKKOTTIL THAMBI.....

Date:13th of August 2023.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
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.	<input type="checkbox"/>

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

Abeen Pulikkottil Thambi
x21224714@student.ncirl.ie:

1 Introduction

My research project submitted as part of the Masters in Fintech completion along with configuration manual . The technologies and hardware configuration of the machine applied in the study are explained in the configuration handbook. Moreover, the research study intends to serve as a resource for scholars by detailing how they might replicate the results of this study.

2 System Specification

2.1 Hardware of the System

- CPU: 8-core multi-core processor. Intel Core i7 architecture is cutting-edge.
- NVIDIA GPU (Graphics Processing Unit) with CUDA and OpenCL support. For machine learning, a GPU with 8GB VRAM is recommended. RTX4050A by NVIDIA
- RAM (memory): DDR4 16GB RAM. Solid State Drive (SSD) storage enables faster data access and model loading. SSD = 512 GB

2.2 Software and Tools

- The most recent version (3.7.4) of Python. NumPy, pandas, matplotlib, Sklearn, and TensorFlow are all Python libraries.
- Machine Learning Frameworks: PyTorch with GPU-enabled versions installed for improved training performance.

3 Data Source

Google Forms- The authorised questionnaires for the study were incorporated in the Google Forms application. This was distributed to participants in order to gather information to use in the study. The collected data was recorded in a Google spreadsheet before being exported as a locked password Microsoft csv file.

Figure 1

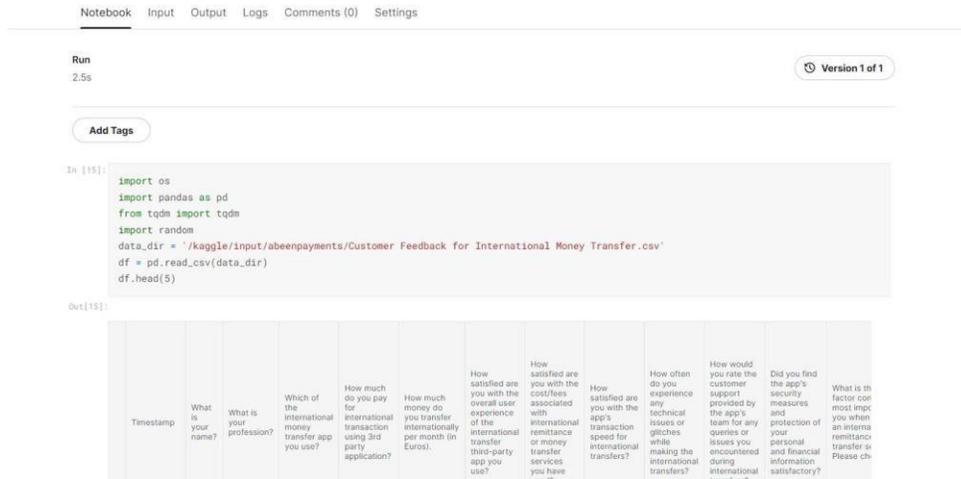
What is the main 3 factor considered most important to you when choosing an international remittance or money transfer service? Please check only 3. *

- Cost/Fees
- Exchange Rates
- Transaction Speed
- Reliability
- Ease of Use
- Security
- Customer Support
- Network Coverage
- Delivery Options
- Other: _____

Jupyter Notebook: Jupyter Notebook is a tool that is a free and free web tool that lets you build and distribute notebooks with live code, formulas, visualisations, and explanatory text. It's frequently used in artificial intelligence for participatory testing, development, and explanation. Here are some important aspects of Jupyter Notebook:

- Jupyter Notebook provides a collaborative setting in which you may run code in tiny pieces known as cells. It makes it simple to play around with code, investigate theories, and see outcomes in the moment.
- Mixed Media: It allows you to combine code, text, computations (using LaTeX), graphics, and visualisations in a single record, which makes it excellent for recording your processes of thought and outcomes.

Figure 2



4. Data cleaning and Preprocessing

The snowball method was utilised to obtain information for this research, which included 75 people who responded and over fifteen surveys. In the snowball randomization study approach, initial subjects who fit specified criteria are picked, and these individuals are then asked to suggest prospective participants that match identical requirements. While this technique is repeated in a chain-like style, the original specimen size is raised. It is typically employed when researching secretive or marginalised societies when normal sample approaches are not possible.

Table 1

S. No	Question	Options
1	What is your profession?	1) Teacher 2) Doctor 3) Nurse 4) Soliciter 5) Others
2	Which of the international money transfer app you use?	1) Wise 2) Revolut 3) Remitly 4) OFX 5) Western Union 6) Others
3	How much do you pay for international transaction using 3rd party application?	1) 20% 2) 40% 3) 50%

		4) 50%+ 5) Others
4	How much money do you transfer internationally per month (in Euros).	
5	How would you rate the customer support provided by the app's team for any queries or issues you encountered during international transfers?	1) 1 (Worst) 2) 2 3) 3 4) 4 5) 5 (Best)
6	How satisfied are you with the cost/fees associated with international remittance or money transfer services you have used?	1) Very Satisfied 2) Satisfied 3) Neutral 4) Dissatisfied 5) Very Dissatisfied
7	How satisfied are you with the app's transaction speed for international transfers?	1) 1 (Worst) 2) 2 3) 3 4) 4 5) 5 (Best)
8	How often do you experience any technical issues or glitches while making the international transfers?	1) Very Often 2) Never
9	Which of the following option do you trust the most?	1) Banks 2) Third party remittance application

Figure 3

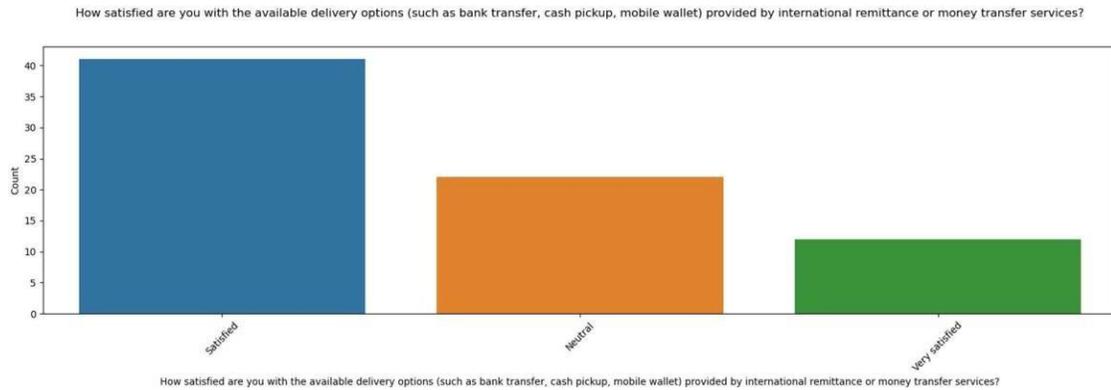
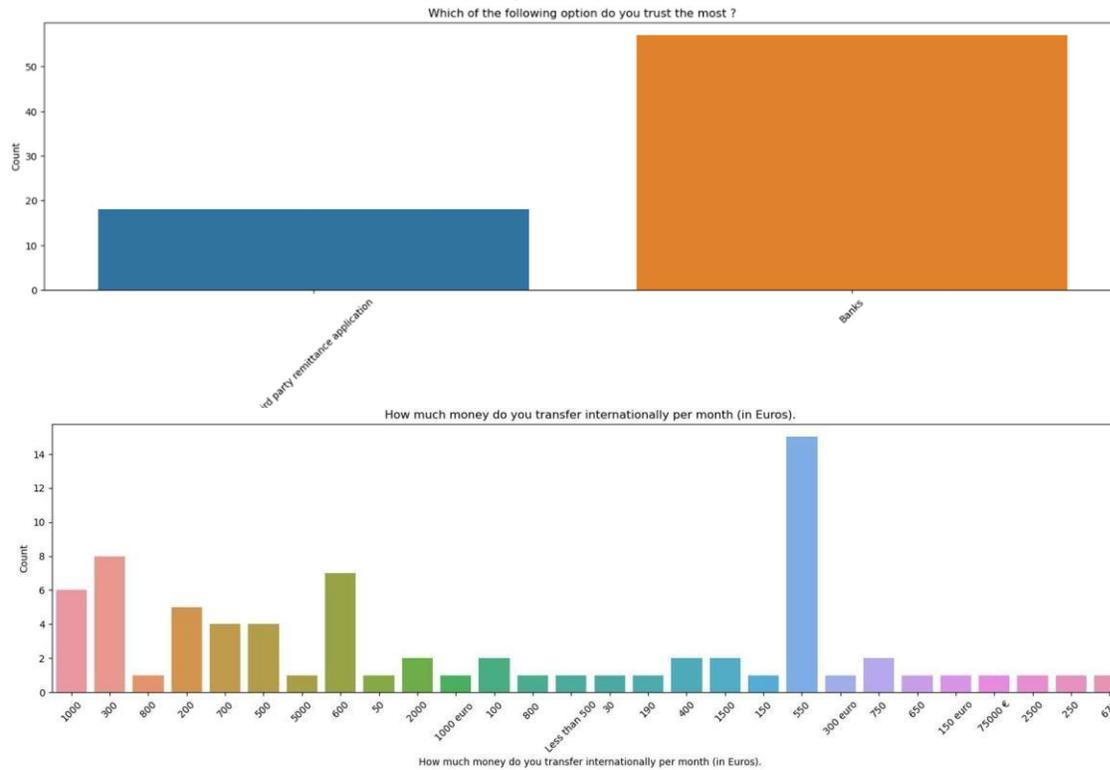


Figure 4 and 5



The technique of turning classification variables (textual names) into numbers is known as label encoded information. Encoding labels is one method for providing numerical data to neural network systems.

Cardinal vs. Hypothetical Values: Encoding of labels may be acceptable for ordinal numbers (in which the categories have an intrinsic order), but it wouldn't necessarily be needed for nominal numbers (in which there is no specified ordering among classes).

Normalisation, commonly referred to as the scaling of features, is the process of converting numerical characteristics to a standardised scale. This guarantees that characteristics with varying sizes or units participate

evenly to the educational process. Normalisation procedures that are commonly used include. Standardisation(Z-Score): Converts characteristics with an average of 0 and a deviation from the mean of 1.

5. Techniques Used

Exploratory Data Analysis

Status in Brief

- Calculate fundamental statistics for numerical columns of data (for example, the mean, the median and standard deviation).
- Compute frequency counts for every group in the categorical columns.

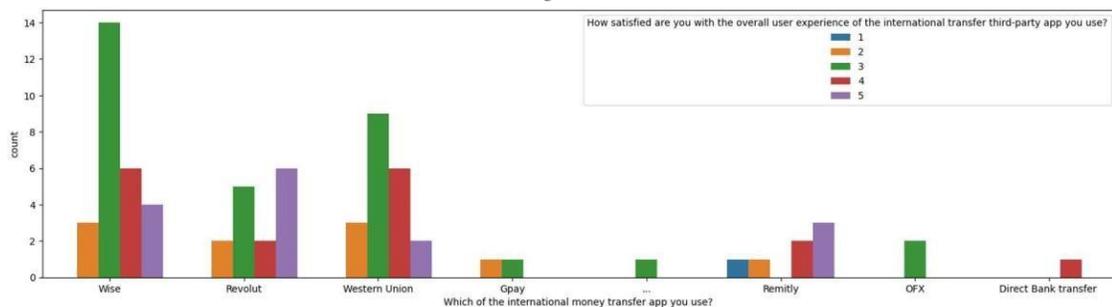
Analysis of Variables

- Analyse the order of replies for every query.
- Employ graphs with bars and pie graphs to depict the spread of categories such as "Profession" and "International Payment Transfers Application".

Bidirectional Evaluation

Investigate the links among elements. To contrast answers across categories, use bar graphs or aggregated bar graphs. Comparing the "Profession" data to the "Money Transfer App" variable, for instance, to determine which applications are preferred among various occupations.

Figure 6



Illustrations

- Develop visualisations that help you comprehend the information better.
- Histograms are may be used to visualise the spread of ongoing information such as "Money Transferred per Month".

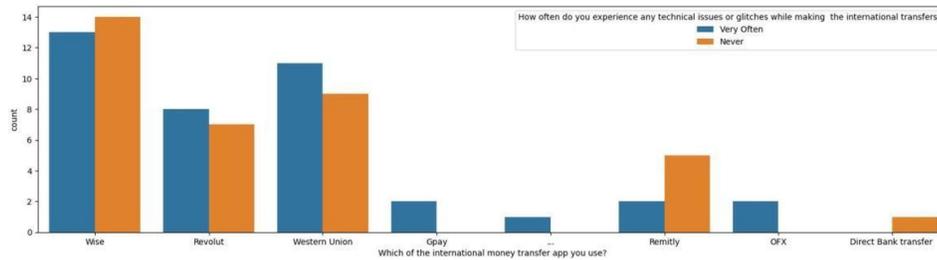


Figure 7

6. Machine Learning

Figure 8

```
X_train, X_test, y_train, y_test = train_test_split(OP,Y,test_size=0.2, random_state=1)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
from sklearn.naive_bayes import GaussianNB
from sklearn import svm
rdf = pd.DataFrame()
models_f1 = []
models_f1.append((RandomForestClassifier(n_estimators=710, criterion='gini', max_depth=None,
                                         min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
                                         max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0,
                                         bootstrap=True, oob_score=False, n_jobs=None, random_state=None,
                                         verbose=0, warm_start=False, class_weight=None)))
models_f1.append((DecisionTreeClassifier(criterion='gini', splitter='best', max_depth=None,
                                         min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
                                         max_features=None, random_state=None,
                                         max_leaf_nodes=None, min_impurity_decrease=0.0,
                                         class_weight=None, ccp_alpha=0.0)))

models_f1.append(GradientBoostingClassifier())
```

Figure 9

```
def visualise(y_true,y_pred,name,model_name,subset,types):
    ConfusionM = confusion_matrix(y_true,y_pred)

    print('*****'+types+'*****')
    print('ROC AUC score:',
          multiclass_roc_auc_score(y_true,y_pred, model_name + '___'+subset))
    print(classification_report(y_true,y_pred,target_names = ['Neutral', 'Satisfied', 'Very satisfied'], digits=4))
    print()
    fig, ax = plt.subplots(figsize=(10,10))
    sns.heatmap(ConfusionM, annot=True,
               fmt='g',xticklabels=le.classes_, yticklabels=le.classes_)
    plt.title('Confusion Matrix for '+model_name)
    plt.savefig('CM_'+model_name + '___'+subset+'.jpg')
    plt.grid()
    plt.show()
    return onevsall(y_true,y_pred,model_name,types)

def run(clf,X_train, y_train,X_test,y_test,subset):
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    y_true = y_test
    name = np.unique(y_train)

    visualise(y_test,y_pred,name,clf.__class__.__name__,subset,'Test')

def RNDf(X,y,subset):
    for clf in tqdm(models_f1):
        print(clf.__class__.__name__)
        X_train, X_test, y_train, y_test = train_test_split( X, y, test_size=0.4)
        run(clf,X_train, y_train,X_test,y_test,subset)
    return rdf
```

i) Random Forest Classifier

The RandomForestClassifier is a collective approach to learning that makes forecasts by combining numerous decision trees. It is very helpful for categorization jobs. The following are the variables that are utilised by the source code snippet:

Estimators: The total amount of plants in a forest in this scenario (710 in this example).

Criterion: The purpose used to assess the worth of a divided substance (in this example, Gini deviation).

Max_depth: The deepest level of each of the branches (None indicates that there isn't a deeprestriction).

Min_samples_split: The lowest amount of specimens needed to divide a node within the network(2).

Min_samples_leaf: The smallest amount of specimens needed at an individual leaf nodes in thisscenario (1).

Max_features: the amount of attributes to examine for determining the appropriate split (sqrtdenotes the cubic root of all of the characteristics).

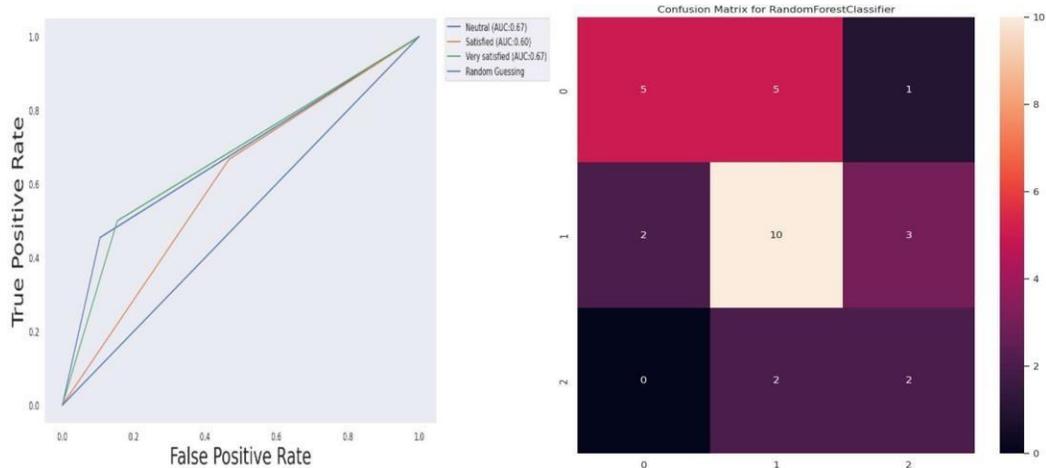


Figure 10

ii) Decision Tree Classifier

To make forecasts, the decision tree classifier constructs just one tree of decisions. It is an essential component of ensemble techniques such as random forest construction. The following are the variables that are utilised by the source code snippet:

- Criterion: the purpose used to assess the worth of a divided substance (in this example, Gini impurity).
- Splitter: The approach used to select the split for the node level (in this example, the optimal split determined by gini impurities).

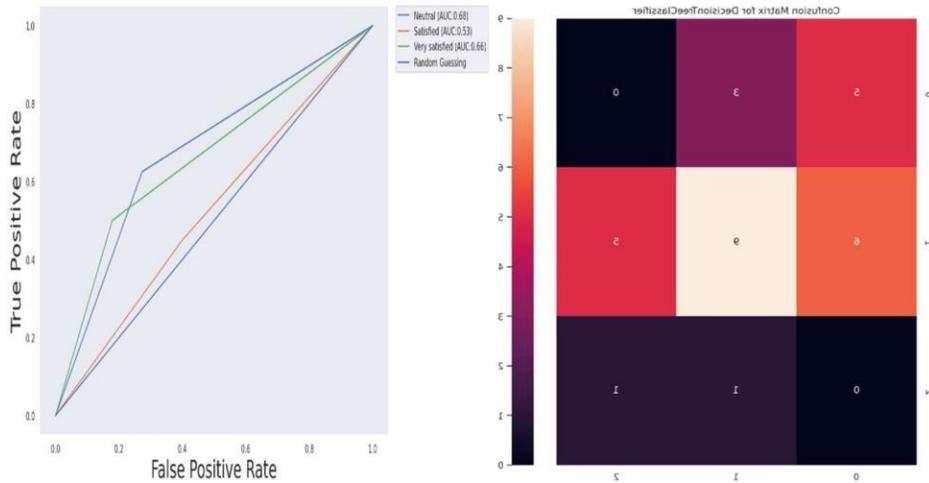


Figure 11

iii) Gradient boost

It is an ensemble approach which brings together the forecasts of numerous novice learners (usually decision trees) in order to produce a powerful model for forecasting. The GradientBoostingClassifier is employed here sans giving any of its hyperparameters directly.

We are adding occurrences of the mentioned classifications to the models_f1 list with this codeline. Given the characteristics presented, such classifiers may be trained and tested on an assortment of data to estimate happiness levels. It's vital to remember that the accuracy of such models is affected by the quality and amount of the data set, as well as hyperparameter values tweaking. Cross-validation is and parametric adjustment should also be carried out to verify that the models generalise effectively to fresh data.

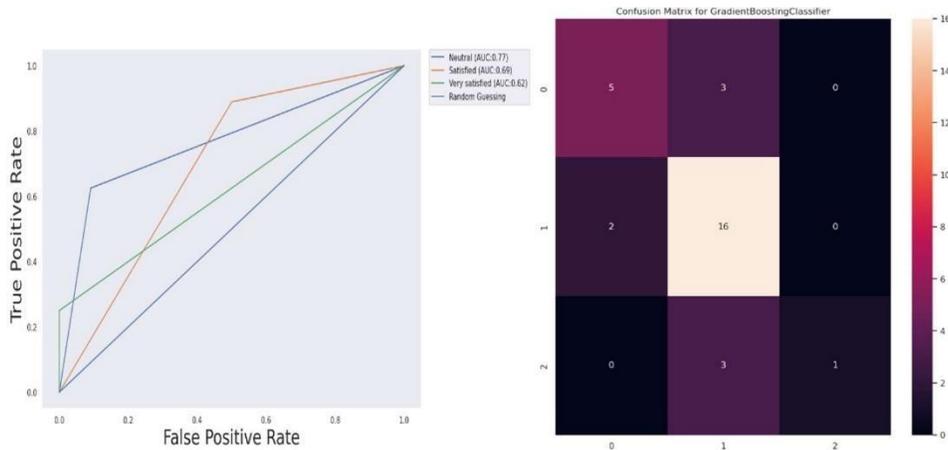


Figure 12

7. Conclusion

This setup manual outlines the primary technologies which were employed in this study. It goes over how the information was gathered, sanitised handled, and analysed. The participants in this research were polled using a methodical interview and an internet survey. This handbook offers details that will assist in the replication of the study outcomes.