

Dataset Link - <https://remittanceprices.worldbank.org/node/188991/download/cb5f8e1938e18be3f54b1ecce6bb4614>

About the dataset

- **id**: Unique identifier for each transaction.
- **period**: The time period when the data was collected (e.g., 2016_2Q).
- **source_code**: Code representing the sender's country.
- **source_name**: Name of the sender's country.
- **source_income**: Income classification of the sender's country (e.g., Lower middle income).
- **destination_code**: Code representing the receiver's country.
- **destination_name**: Name of the receiver's country.
- **destination_region**: The region where the receiver's country is located.
- **destination_income**: Income classification of the receiver's country.
- **destination_lending**: Lending classification of the receiver's country (e.g., IBRD, IDA).
- **firm**: The company facilitating the transaction.
- **firm_type**: Type of the firm (e.g., Bank, MTO).
- **payment instrument**: Method used for the payment (e.g., Cash, Bank transfer).
- **speed actual**: Actual time taken to complete the transaction.
- **cc1 lcu amount**: The amount of money sent in local currency units (LCU).
- **cc1 lcu fee**: The fee charged for the transaction in LCU.
- **cc1 fx margin**: The foreign exchange margin for the transaction.
- **cc1 lcu fx rate**: Rate at which the amount is being transferred.
- **cc1 total cost %**: The total cost of the transaction as a percentage.
- **transparent**: Whether the transaction is considered transparent.
- **receiving network coverage**: The network coverage at the receiving end (e.g., High, Low).
- **pickup location**: The location where the receiver picks up the money.
- **pickup method**: The method used to pick up the money (e.g., Cash, Bank account).
- **date**: The date when the transaction was processed.
- **corridor**: The remittance corridor (e.g., AGONAM).

I am filtering the dataset where the destination country is India, meaning the money is being sent to India from any other country, and using the dataset to predict the best firm to transfer the money.

Columns to consider for building the model because these were most relevant for the prediction of the firm.

- cc1 lcu amount
- cc1 lcu fx rate
- source_name
- firm (target)

I am performing one-hot encoding on the categorical columns and building a decision tree model using an 80-20 Split. I have chosen a decision tree algorithm because there were outliers in the dataset.

Accuracy on the training dataset: 0.6959058733499139

Accuracy on the testing dataset: 0.70734506503443

Recall: 0.70734506503443

The dataset is highly imbalanced with 160 classes, the accuracy and recall scores are good.

Money Transfer Firm Predictor

Send Money to India

Amount to Transfer

735

Exchange Rate

15

Source Country

United Arab Emirates

Speed of Transfer

Less than one hour

Predict Firm

	Amount to Transfer	Exchange Rate	source_name_Australia	source_name_Bahrain	source_name_Ca
	0	735	15	0	0

The predicted firm for this transaction is: Al Fardan Exchange

I have created a web application for the firm predictor using Streamlit.