Income & Expense Prediction Using Machine Learning

Introduction:

Financial transaction data provides deep insights into customer behavior, spending patterns, and income flow.
In this project, we analyze real ANZ bank transaction data and build predictive machine learning models to estimate:

- **Total monthly income of customers**
- **Total monthly expenses of customers**

Dataset Description:

Each row represents a unique transaction performed by a user.

Key Columns Used

- customer_id: Unique identifier for each user
- movement: credit/debit (used to calculate income/expenses)
- amount: transaction value
- txn_description: transaction type (POS, salary, payments, etc.)
- balance: account balance
- date: transaction timestamp

Objective:

The two major predictive tasks:

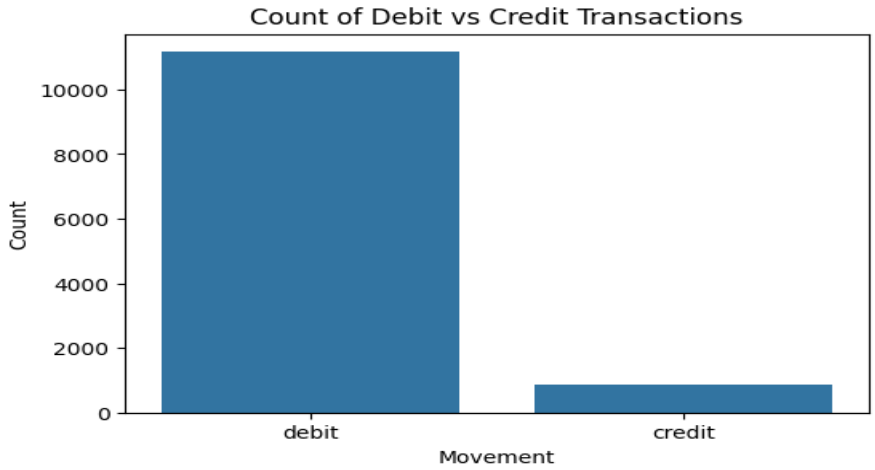
1. Predict Total Income per User

Income = sum of all credit transactions for that user.

2. Predict Total Expenses per User

Expenses = sum of all debit transactions for that user.

These targets are derived mathematically from the dataset without any external data.



Feature Engineering:

- **Total transactions**
- **Debit transaction count**
- **Credit transaction count**
- **Average transaction amount**
- **Maximum transaction amount**
- **Minimum transaction amount**

These features represent customer financial behavior and were merged with targets (total income, total expenses) to create a machine-learning-ready dataset.

Machine Learning Models:

Two types of models were used:

1. Linear Regression

2. Random Forest Regressor

Separate models were trained for:

- Income prediction
- Expense prediction

Train-test split:

- 80% training
- 20% testing

Model Evaluation Metrics:

The models were evaluated using:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**
- **R^2 Score**

Where:

- Lower MAE/RMSE = better accuracy
- Higher R^2 = better model fit

Results:

Expense Prediction (Good Performance)

Metric	Value
MAE	~934 AUD
RMSE	~1358 AUD
R^2	**0.65**

Interpretation

- Model predicts expenses with **good accuracy**
- Strong, consistent features = good model performance

Income Prediction (Poor Performance)

Metric	Value
MAE	~10,588 AUD
RMSE	~12,053 AUD
R^2	**-2.54**

Interpretation

- Model performs poorly on income

- Income cannot be predicted accurately with current features

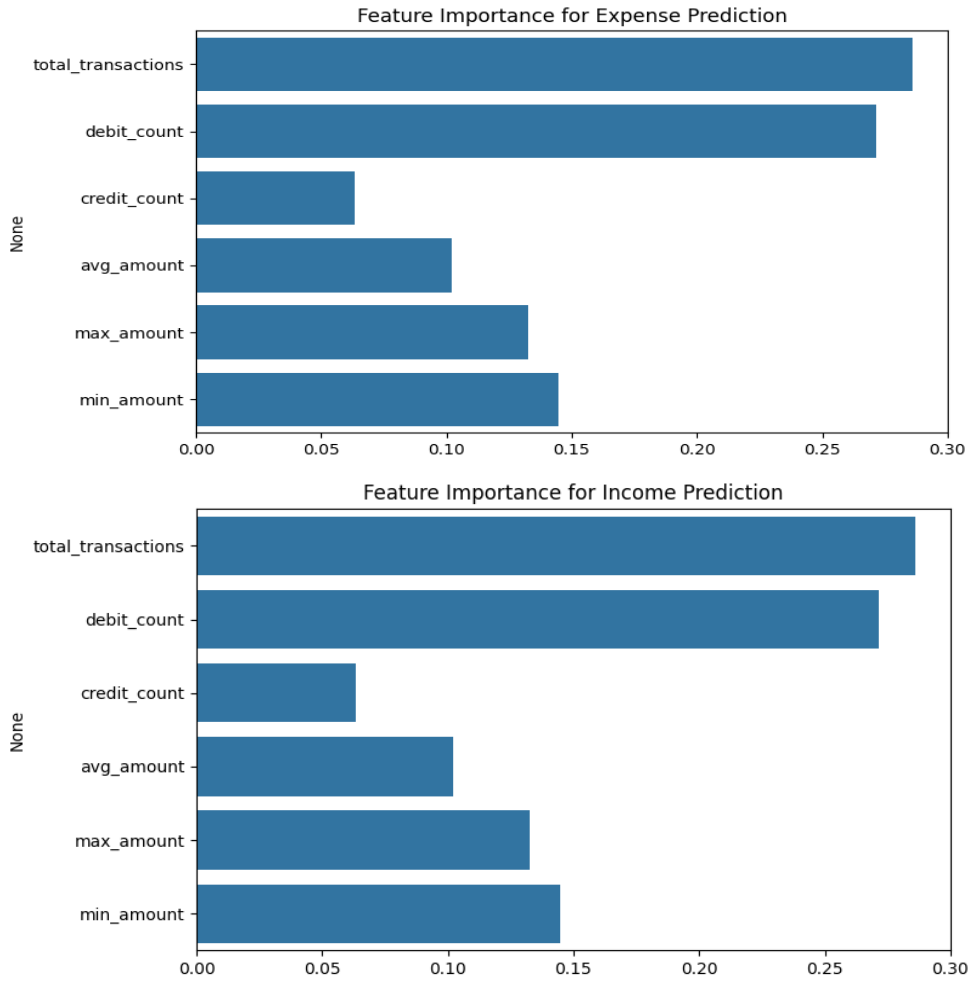
---

Feature Importance Analysis:

Expense Model – Useful Features

- total_transactions
- debit_count
- avg_amount
- max_amount

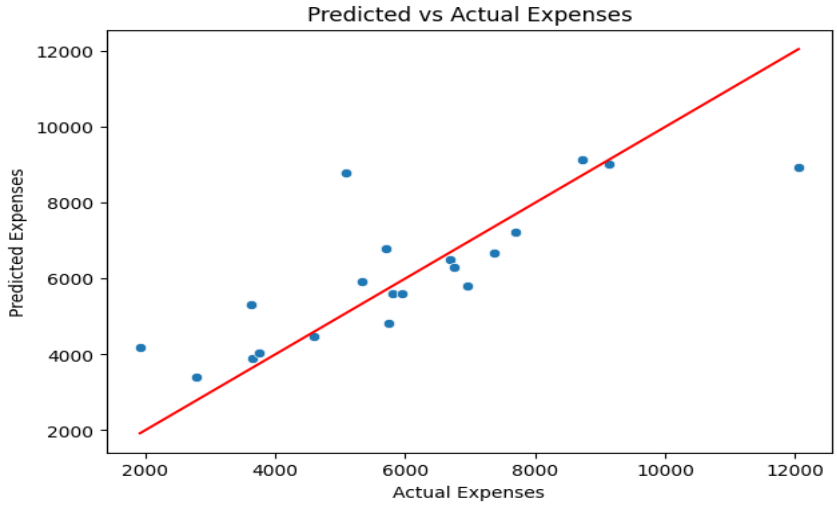
These features strongly correlate with customer spending behavior.



Predicted vs Actual Analysis:

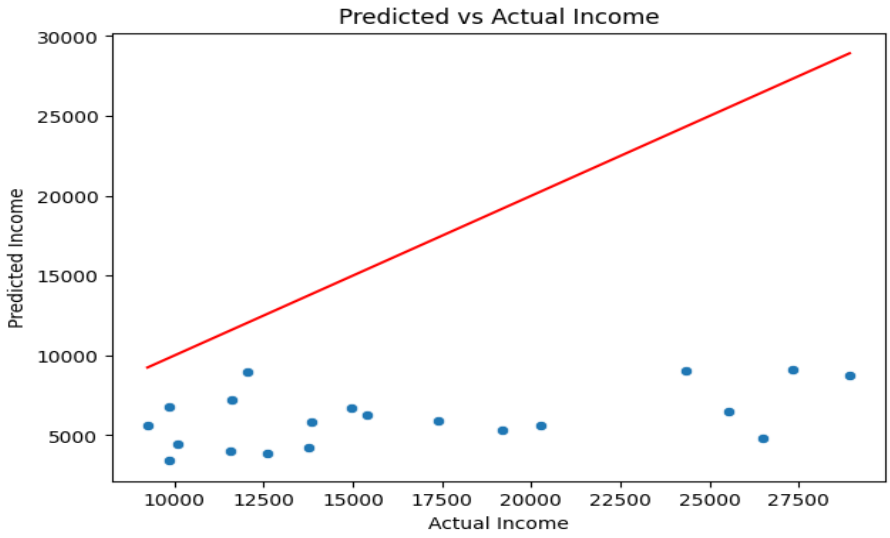
Expenses

- Scatter plot shows points close to the diagonal line
- Indicates stable and accurate model



Income

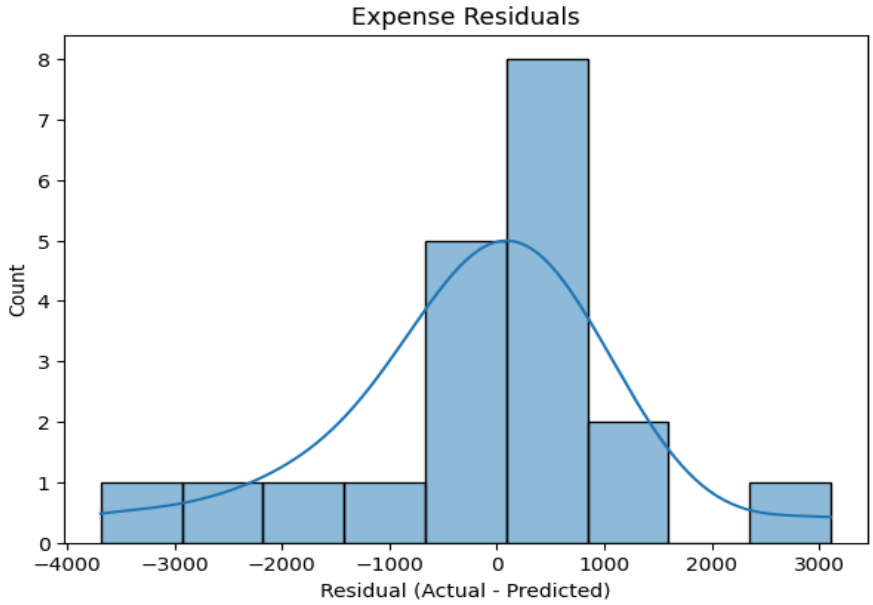
- Points are scattered far from the line
- Model cannot learn consistent income patterns



Residual Analysis:

Expenses

- Residuals centered around 0
- Model errors are small and balanced



Income

- Residuals show wide spread
- Indicates unpredictable patterns
- Confirms poor model fit