Smart Personal Expense Tracker with Category Prediction

1. Project Overview

The Smart Personal Expense Tracker with Category Prediction is a data-driven application that simplifies financial management by combining expense tracking with machine learning. By analyzing expense descriptions, the system intelligently predicts categories such as food, travel, shopping, and bills. It reduces manual input, enhances budgeting accuracy, and offers real-time insights through a user-friendly dashboard.

2. Objectives

- Fetch live transaction data using Plaid API
- Store transactions securely in MySQL
- Automatically classify transactions into categories
- Analyze income and expenses
- Predict future spending using ML
- Display insights using an interactive dashboard

3. Technology Stack

- Programming Language: Python
- Database: MySQL
- API: Plaid (Sandbox)
- Machine Learning: Scikit-learn (Linear Regression)
- Dashboard: Streamlit
- Visualization: Plotly, Altair, Matplotlib
- Tools: Pandas, NumPy, Joblib

4. System Architecture

1. Plaid API fetches live transactions
2. Python script processes and categorizes data
3. Transactions stored in MySQL database
4. ML model predicts spending totals
5. Streamlit dashboard visualizes real-time insights

5. Database Design

Table: transactions

- transaction_id (Primary Key)
- user_id
- product_id
- quantity
- price
- total
- transaction_time
- category_id
- description

Table: Category

- Category id
- Category name
- Description

6. Data Collection (Live Fetch)

- Transactions are fetched using Plaid Sandbox API
- Data includes price, date, and transaction name
- Python script runs continuously and inserts data every 60 seconds

7. Category Prediction Logic

Expenses are categorized based on transaction amount:

- Snacks (≤ ₹50)
- Lunch (₹51 – ₹300)
- Shopping (₹301 – ₹1500)
- Electronics (> ₹1500)
- Income (positive amount)

This logic ensures meaningful real-world categorization.

8. Machine Learning Model

- Model Used: Linear Regression
- Features: quantity, price, transaction hour, day of week
- Target: total transaction amount
- Training Data: Historical MySQL transaction data
- Model Storage: Saved using Joblib

9. Dashboard Features

- Total Income, Expense, Net Balance KPIs
- Daily and Monthly Expense Trends
- Category-wise Expense Analysis
- Top 10 Transactions
- Income vs Expense Pie Chart
- Heatmap (Category vs Time)
- Budget Tracking
- 7-Day Spending Forecast
- Auto-refresh every 60 seconds

10. Smart Alerts

- Daily overspending alert
- Financial activity status (Good savings / Overspending)

11. Results & Insights

- Users can clearly understand spending habits
- Automatic categorization saves manual effort
- ML-based prediction helps future planning
- Dashboard provides real-time decision support

12. Machine Learning Predication

```
sample data :
   transaction_id                          user_id  ... description  category
0               1  PnRgWpxLBXsJ55WmGenqiNgqDlnXAdfoMLAPX  ...        None    Income
1               2  NkVpN1BLrbUK55yJrnoAsywKJ7zX8ZCyD8Q4y  ...        None    Income
2               3  eN4KXVGRmQTyjjZD4nowUKZERd8ly3Trjvm4E  ...        None    Income
3               4  jQ3x5V9R8Xc5AAwn8dbKIwDB4yLjlEI6XKaVZ  ...        None    Income
4               5  7zvwbAxg93TZEEDnw58pt7yoZ9LXgrCd9bxo1  ...        None    Income

[5 rows x 10 columns]
   quantity   price  transaction_hour  transaction_dayofweek
0         1    25.0                16                      5
1         1     5.4                16                      5
2         1  5850.0                16                      5
3         1  1000.0                16                      5
4         1    78.5                16                      5
0      25.0
1       5.4
2    5850.0
3    1000.0
4      78.5
Name: total, dtype: float64

Model performance
Mean Squared Error : 1.618013670037092e-24
mean absolute error : 9.912071163853398e-13
R2 Score : 1.0

 Model trained successfully
Model saved at: c:\Aparaitech\project_2_expense\data\transaction_model.pkl
```

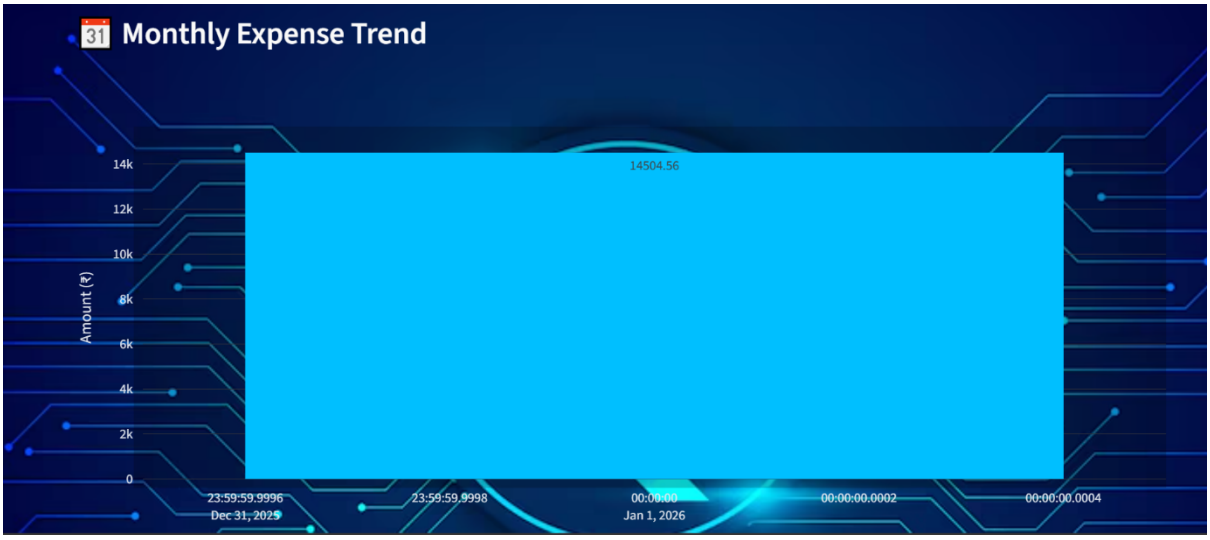
13 . DashBoard Output

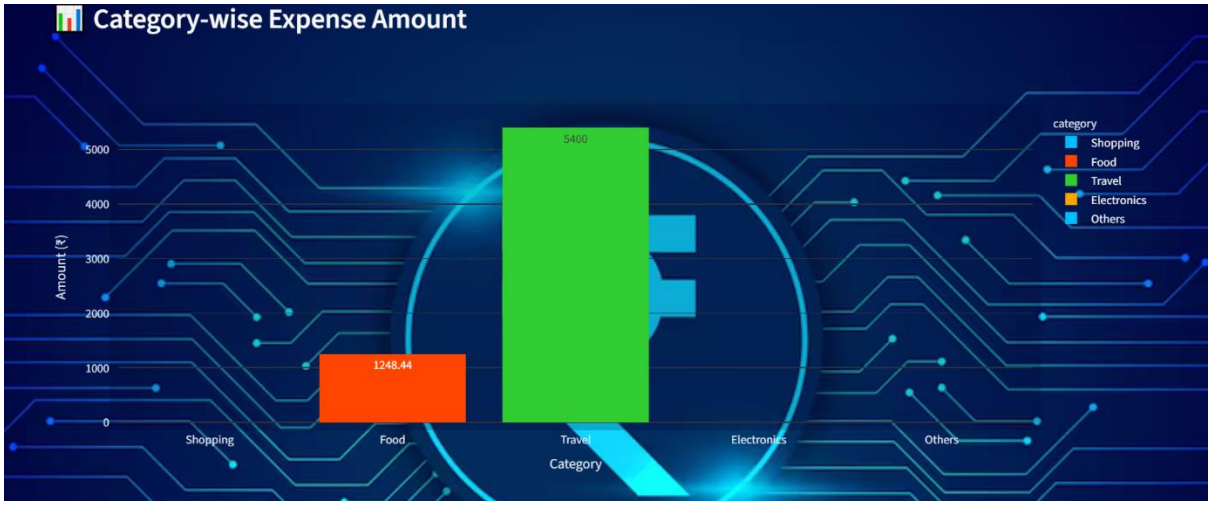
Top 10 Transactions

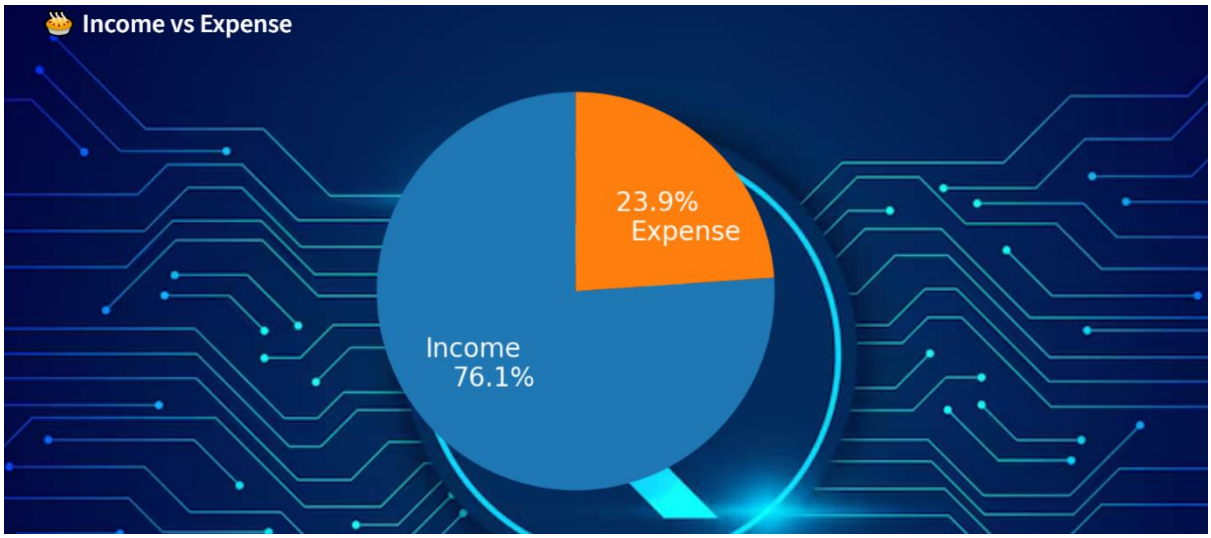
transac...	quantity	price	total	transac...	product...	category	month	hour
3	1	5850	5850	17/01/202...	0	Income	2026-01	16
13	1	5850	5850	19/01/202...	0	Income	2026-01	10
4	1	1000	1000	17/01/202...	0	Income	2026-01	16
14	1	1000	1000	19/01/202...	0	Income	2026-01	10
23	1	1000	1000	19/01/202...	1	Income	2026-01	16
48	1	1000	1000	13/01/202...	1	Income	2026-01	0
43	1	1000	1000	19/01/202...	1	Income	2026-01	17
38	1	1000	1000	19/01/202...	1	Income	2026-01	17
28	1	1000	1000	19/01/202...	1	Income	2026-01	17
33	1	1000	1000	19/01/202...	1	Income	2026-01	17

Hourly Spending Pattern

Monthly Expense Trend




Category-wise Expense Amount
category
Shopping
Food
Travel
Electronics
Others
5400
1248.44
Amount (₹)
5000
4000
3000
2000
1000
0
Shopping
Food
Travel
Electronics
Others
Category

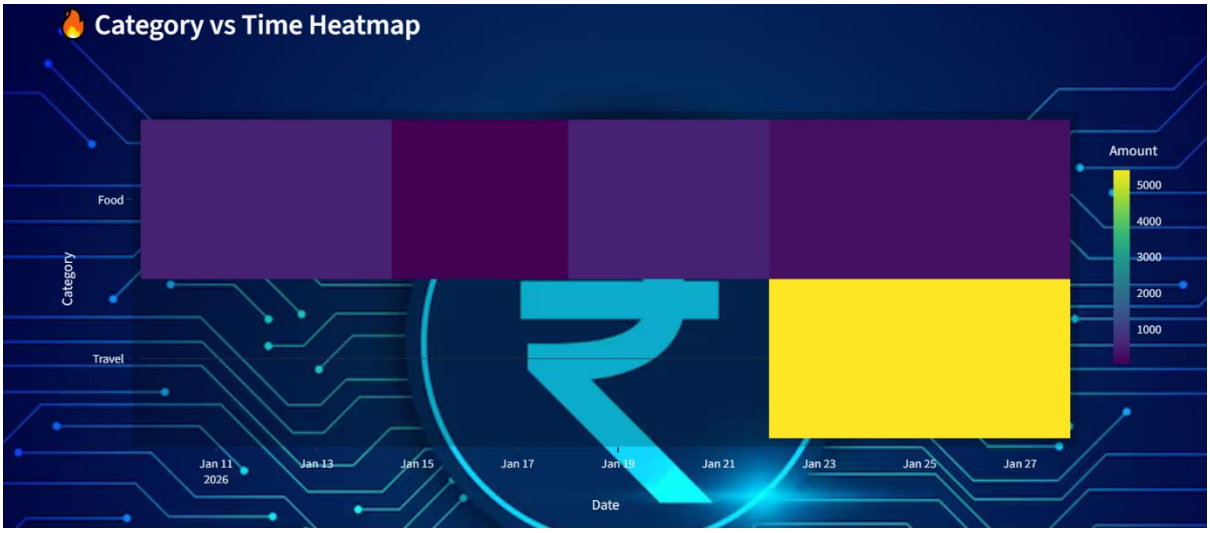

Income vs Expense
23.9%
Expense
Income
76.1%


Cumulative Balance
Cumulative Balance
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
−2,000
Mon 12
Tue 13
Wed 14
Thu 15
Fri 16
Sat 17
Jan 18
Mon 19
Tue 20
Wed 21
Thu 22
Fri 23
Sat 24
Jan 25
Date

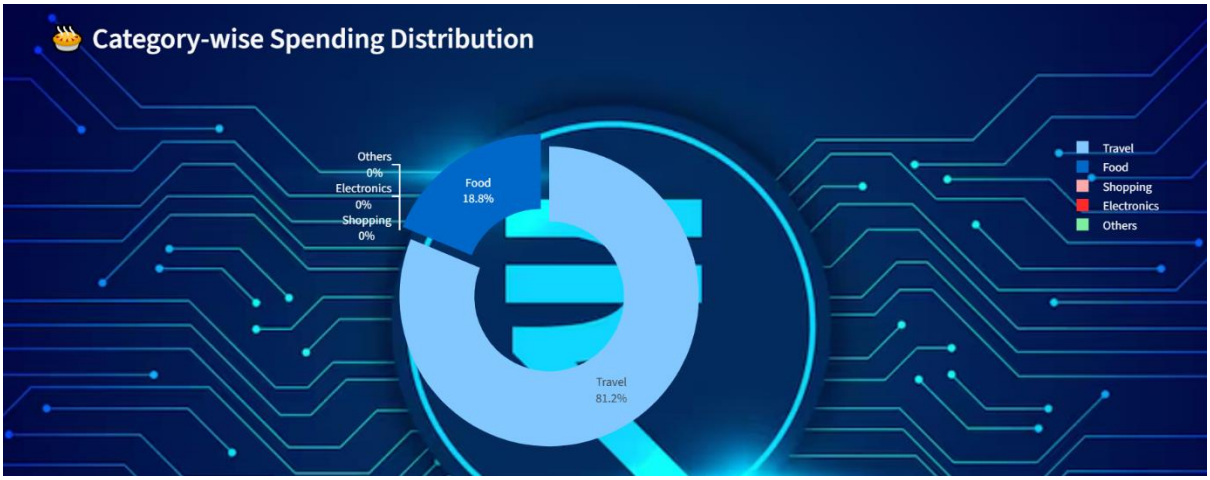
Monthly Budget Tracking

category	Spent	Budget	Remaining
Food	1248.44	1498.128	249.68799999999987
Travel	5400	6480	1080

Category vs Time Heatmap



Category-wise Spending Distribution




Transaction Amount Distribution
Number of Transactions
14
12
10
8
6
4
2
0
−2000
−1000
0
1000
2000
3000
4000
5000


7-Day Spending Forecast
Predicted Spending
400
350
300
250
200
150
100
50
0
Mon 26
12 PM
Tue 27
12 PM
Wed 28
12 PM
Thu 29
12 PM
Fri 30
12 PM
Sat 31
12 PM
February
Date
Auto refresh every 60 seconds

14. Future Enhancements

- User authentication
- Advanced ML models
- Mobile application
- Cloud deployment
- Personalized budget limits

15. References

- Plaid API Documentation
- Streamlit Documentation
- Scikit-learn Documentation

16. Conclusion

The Smart Personal Expense Tracker successfully integrates live data, analytics, and machine learning into a single system. It provides valuable insights for personal finance management and demonstrates practical application of data analytics skills.