IMT 598 - Finance Fraud Detection

Group 3: Zhitong (Mia) Xie, Ajinkya Sheth, Xiaohua Shi

AGENDA

1. INTRODUCTION
2. MACHINE LEARNING ON AZURE
3. DEMO
4. LIMITATIONS

1.
Introduction

Industry

Bank/Financial Industry

- Brick-and-mortar → Online Platform
- Utilized big data to expand core business model
  - Expand customer segment from B2C to B2B (billing platform)
  - Recommendation Engine (personalized service and products)

Business Problem

Freya Group

- Top US-based financial services provider
  - Key service: Mobile Banking
- Current Digital Transformation:
  - On-premise server
  - Regression model to detect fraud transaction

Business Problem

Lack of Efficient Model for Fraud Detection!

1,048,576

Transactions in our dataset

10.13%

Are fraud transactions

0.195%

Fraud transactions are successfully identified!

Business Model Change

Going Cloud!

- Customer: seamless experience
- Company: reduce cost and reliance on hardware; increase flexibility and scalability

Data Provider!

- Improving algorithm
- Bring insights on customer pattern

B2B to B2C!

- Expand to health-care industry
- Expand to insurance industry

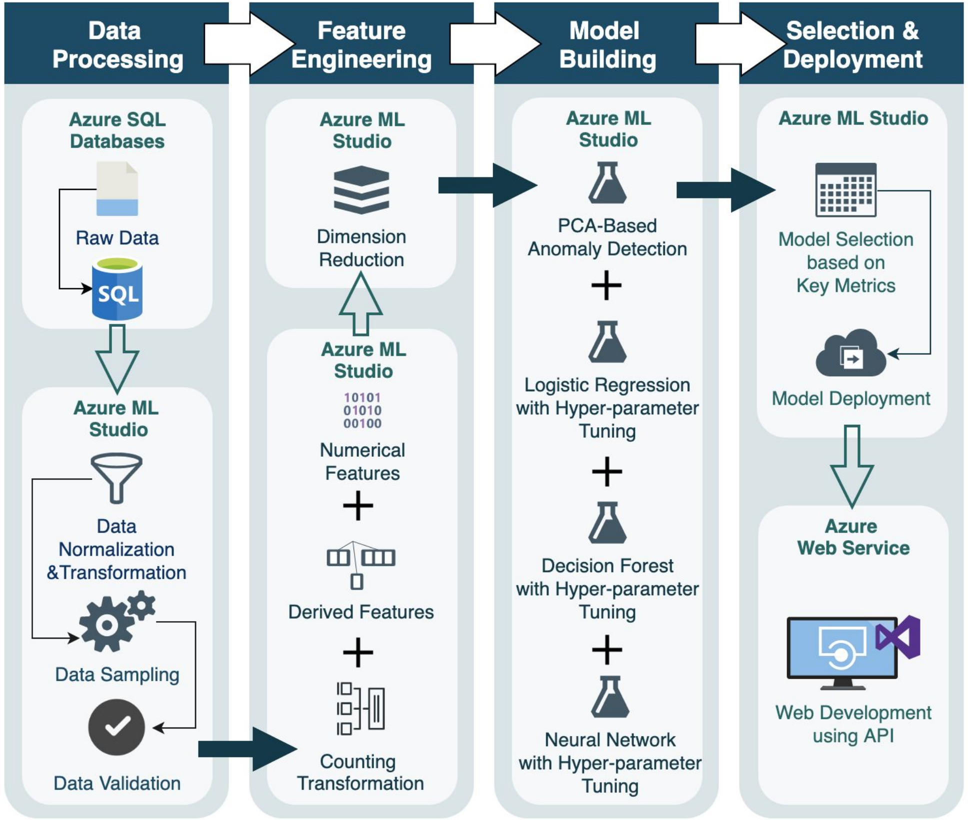
Improving and Expanding!

2.
Solution

Solution

Azure

- Azure SQL Server
- Azure ML Studio
- Azure Web Service



Solution

Raw Data

- Step: hashed time stamp
- Type: [CASH-IN, CASH-OUT, DEBIT, PAYMENT, TRANSFER]
- Amount: amount of the transaction in local currency.
- nameOrig: customer who started the transaction
- nameDest: customer who is the recipient of the transaction
- oldbalanceOrig / oldbalanceDest: initial balance before the transaction for Orig / Dest
- newbalanceOrig / newbalanceDest - initial balance before the transaction for Orig / Dest

Solution – 1st Step Data Processing

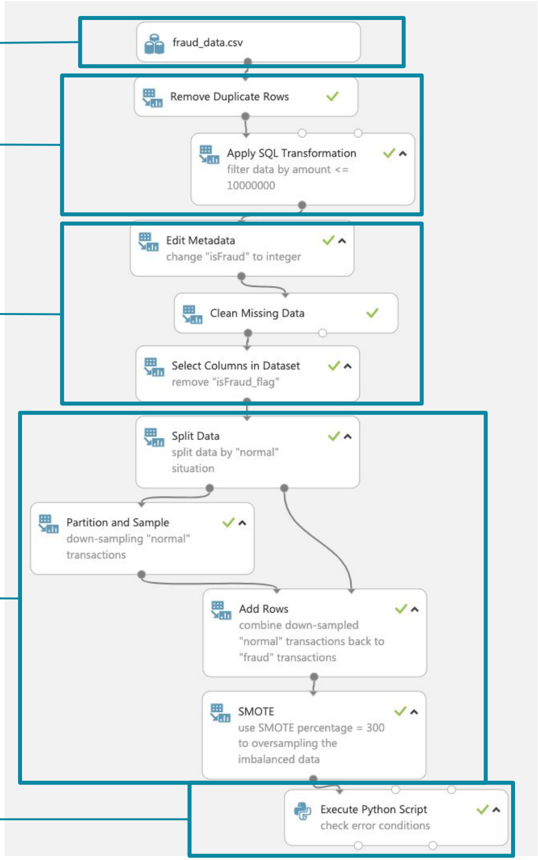
Data Ingestion

Data Transformation

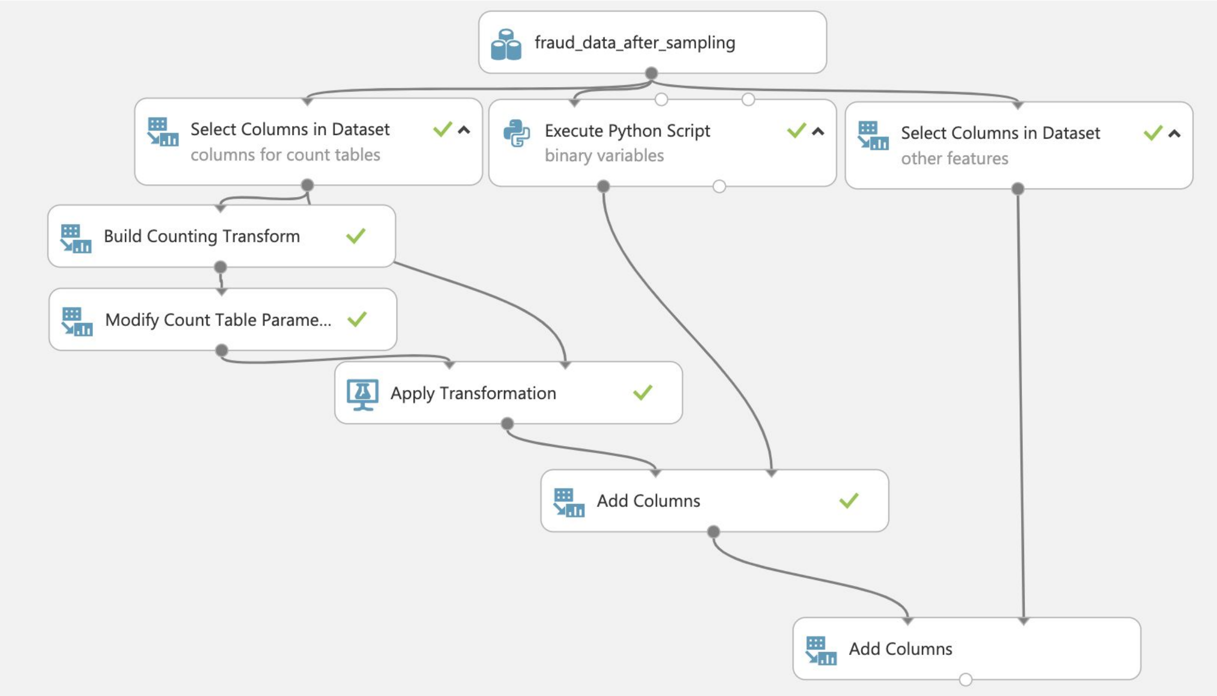
Data Normalization

Data Sampling

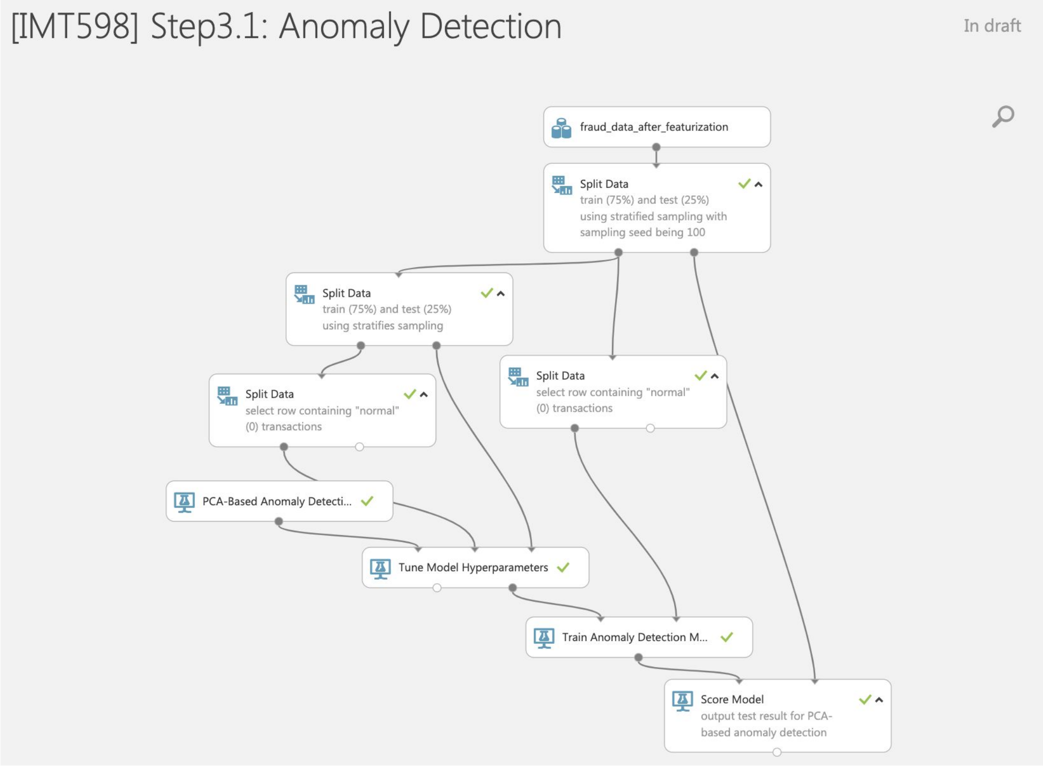
Data Validation



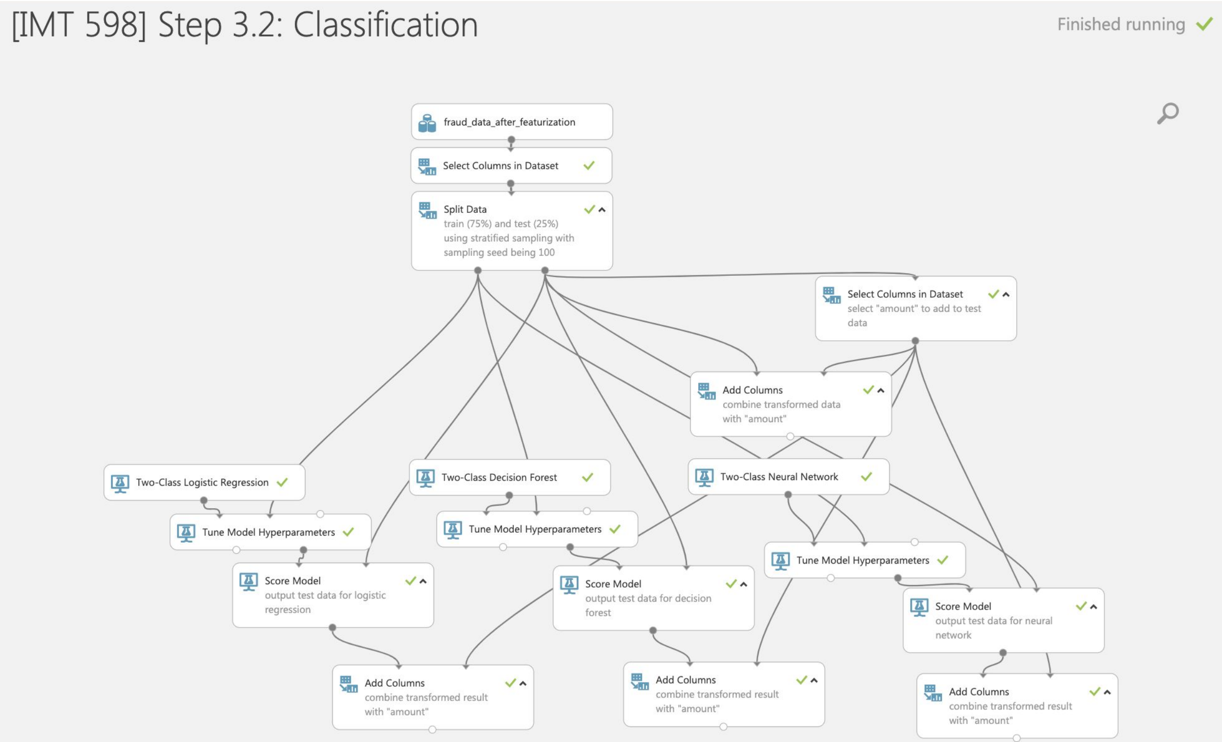
Solution - 2nd Step Feature Engineering



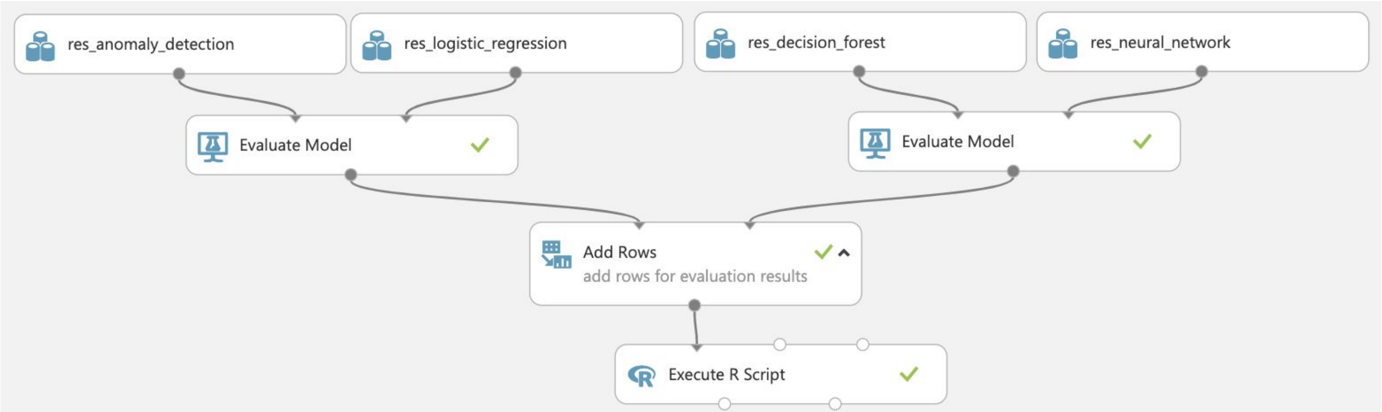
Solution – 3rd Step Model Building



Solution – 3rd Step Model Building



Solution – 4th Step Model Selection



Solution – 4th Step Model Selection

Algorithm	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
Anomaly Detection	0.417036	0.071811	0.721174	0.130616	0.660331	0.706956	-208.776221
Logistic Regression	0.983061	0.972701	0.741873	0.841749	0.991722	0.05444	76.222472
Decision Forest	0.99583	0.967829	0.963351	0.965585	0.998512	0.015648	93.165659
Neural Network	0.98457	0.960463	0.777913	0.859603	0.994018	0.042999	81.219275

Solution – 4th Step Model Selection

Method	ADR	VDR	AFPR
Anomaly Detection	72.12%	81.38%	92.82%
Logistic Regression	74.19%	96.69%	2.73%
Decision Forest	96.34%	99.62%	3.23%
Neural Network	77.79%	97.75%	3.95%
Original Method	0.19%	0.65%	0.00%

Note:
ADR (Fraud Account Detection Rate): The percentage of detected fraud accounts in all fraud accounts.
VDR (Value Detection Rate): The percentage of monetary savings, assuming the current fraud transaction triggered a blocking action on subsequent transactions, over all fraud losses.
AFPR (Account False Positive Ratio): The ratio of detected false positive accounts over detected fraud accounts.

3.

Demo

Web App

- Python based Web App
- Front-end framework : Bootstrap
- Back-end framework : Flask
- Hosted on Linux VM
- URL : `fraudwatch.azurewebsites.net/`

Web Page - Blank

Fraud Watch

Transaction Type:

PAYMENT TRANSFER CASH_IN CASH_OUT DEBIT

Amount:

$

Name Origin:

Transaction Origin

Old Balance Origin:

0

Name Dest:

Transaction Destination

Old Balance Destination:

0

Submit

Web Page #1 – Input

Fraud Watch

Transaction Type:

○ PAYMENT ◉ TRANSFER ○ CASH_IN ○ CASH_OUT ○ DEBIT

Amount:

21312

Name Origin:

C0980

Old Balance Origin:

0

Name Dest:

C9809

Old Balance Destination:

0

Submit

Web Page #1 – Output

Fraud Watch

Transaction Type:

○ PAYMENT ○ TRANSFER ○ CASH_IN ○ CASH_OUT ○ DEBIT

Amount:

$

Name Origin:

Transaction Origin

Old Balance Origin:

0

Name Dest:

Transaction Destination

Old Balance Destination:

0

Submit

Success! {'Fraud Probability': '6.5%'}

Web Page #2 - Input

Fraud Watch

Transaction Type:

○ PAYMENT ◉ TRANSFER ○ CASH_IN ○ CASH_OUT ○ DEBIT

Amount:

181

Name Origin:

C1305486145

Old Balance Origin:

181

Name Dest:

C553264065

Old Balance Destination:

0

Submit

Web Page #2 - Output

Fraud Watch

Transaction Type:

◯ PAYMENT ◯ TRANSFER ◯ CASH_IN ◯ CASH_OUT ◯ DEBIT

Amount:

$

Name Origin:

Transaction Origin

Old Balance Origin:

0

Name Dest:

Transaction Destination

Old Balance Destination:

0

Submit

Success! {'Fraud Probability': '73.78%'}

4.

Constraints and Limitations

Model & Data Constraints

- Lack of time series analysis
  - Due to lack of temporal data
- Only two kinds of Destination Bank Account
  - Private/Customer & Merchant account
- Only one type of Origin Bank Account
  - Private/Customer

Cloud Service Limitations

SaaS Limitations

- Free Tier ML Studio
- Included transactions (per month) : 1,000
- Included compute hours (per month) : 2

PaaS Limitations

- Free Tier Linux VM
- RAM 1 GB, Storage 1 GB
- 60 min/day compute time
- Insecure Data Transfer (Lack of SSL Certificate)

Unanswered Business Questions

- How much time a bank should maintain the data for getting better prediction?
- What is the best probability threshold to flag a transaction as fraud?
  - Currently, we are using 50%

Future Work

- Improve the rule for triggering block system

  [0, 0.5]: GOOD

  [0.5, 0.7]: Contact customer for verification

  [0.7, 1]: BAD

THANKS!

Any questions?

You can find me at:

ajinkya@uw.edu, xzhitong@uw.edu, shi249@uw.edu