
# Real-time credit card fraud detection with Microsoft Azure

Team Members : **Seoyeon Choi, Paniz Ojaghi, Vinay Ravi, Matthew Schoenhofer**

# CREDIT CARD FRAUD STATISTICS

## Global Losses

**$408.5 billion**

In expected card fraud global losses over the next decade (1, Federal Trade Commission) (2, dataprot.net)

## Seniors

**$3 billion**

Lost by seniors to financial scams in 2020 (3, Nilsonreport.com)

## Data Breaches

**135%**

Increase in card data available on the Dark web in the last year (4, Intsights.com)

*Nilson Report. (2021, December)*
*IntSights Cyber Intelligence. (2018, July 11)*

# VICTIM DEMOGRAPHICS (5)

**Reported and documented fraud:**

20-29 years old **44%**

70-79 years old **20%**

**Median LOSS**

325$
20-29 YEARS OLD

635$
70-79 YEARS OLD

1300$
80+ YEARS OLD

*(Comparitech.com, 2022)*

# WHY ELDERS ARE THE PERFECT TARGETS?

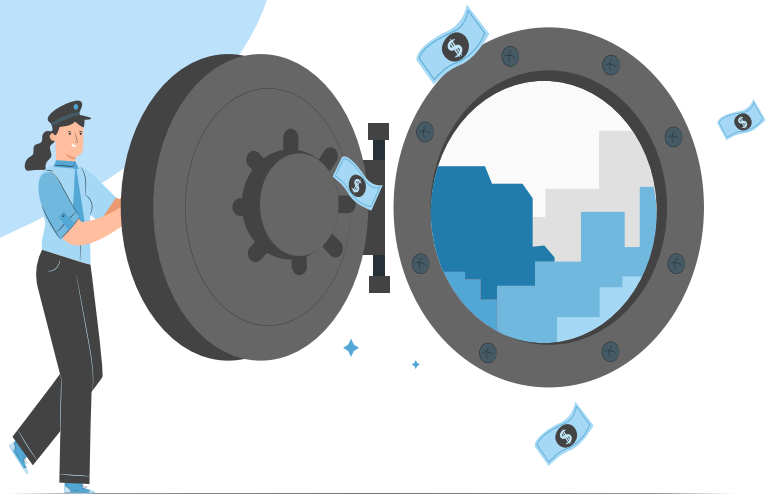
- Largely retired and have a lifetime worth of savings in their bank accounts
- Less likely to have strong knowledge of technology and how to use it
- Sharing personal information or verification online is becoming more common
- Visiting physical branches is becoming less common

# PRESENTATION OVERVIEW

- THE PROBLEM
- THE PROPOSED SOLUTION
- TASKS AND IMPLEMENTATION
- IMPACT AND CHANGE MEASUREMENT
- PRODUCT VIABILITY AND SUSTAINABILITY
- ETHICAL CONSIDERATIONS
- FUTURE RESEARCH

# BENEFITS OF PROBLEM ANALYSIS AND SOLUTION

- Greater knowledge of senior financial fraud risks, causes and importance
- Better understanding of proposed solution process, techniques and implementation
- Stronger appreciation for proposed solutions to impact and change millions of lives for the better
- Recognize the proposed solution is in strong accordance with Microsoft’s Corporate Social Responsibility (CSR)



# OUR OBJECTIVE

## Research Database

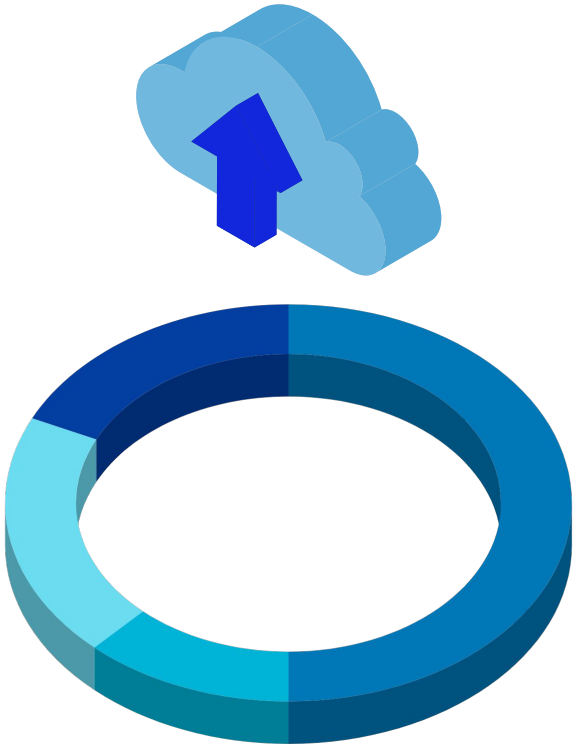
Collection of more than 250,000 transactions, with 492 frauds.(6)

## Build and Train

Using Azure PaaS, machines will learn to detect patterns of fraudulent commerce

## Test and Deploy

Azure container instance for real time interface will help us to test and deploy the model



*(kaggle.com, 2018)*

## Accessible

No need for technical knowledge!
positive societal view of financial institutions

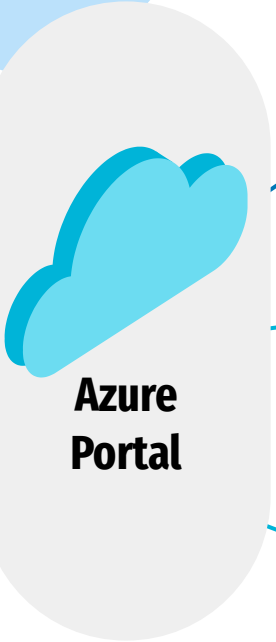
## Continuous stream

Stay up-to-date whenever there are new transactions associated with potential fraud activity

## Accuracy

Better predictions in classifications with high accuracy by AI
More cost effective comparing to old methods

# BASICS




01 Create resource group and cost alert, including steps such as choosing region and subscription

Resource Group

02 Create workspace and compute resources (CPU vs GPU, instances vs clusters, number of nodes, etc.)

ML Studio

03 Azure event grid, Office 365 API connections are used to connect Logic Apps to SaaS services

Azure Logic APPS

# TRAIN, TEST, REGISTER(6)

**Meanwhile in notebook :**

```
import pandas as pd
dataset = pd.read_csv('creditcard.csv')
dataset.drop('Time', axis = 1, inplace = True)
val = dataset['Amount'].values
dataset['Amount'] = StandardScaler().fit_transform(val.reshape(-1, 1))
print(dataset.head(3))
```

```
         V8        V9       V10  ...       V21       V22       V23       V24  \
0  0.098698  0.363787  0.090794  ... -0.018307  0.277838 -0.110474  0.066928
1  0.085102 -0.255425 -0.166974  ... -0.225775 -0.638672  0.101288 -0.339846
2  0.247676 -1.514654  0.207643  ...  0.247998  0.771679  0.909412 -0.689281

        V25       V26       V27       V28    Amount  Class
0  0.128539 -0.189115  0.133558 -0.021053  0.244964  False
1  0.167170  0.125895 -0.008983  0.014724 -0.342475  False
2 -0.327642 -0.139097 -0.055353 -0.059752  1.160686  False
```

**Load the dataset**

**Direct import to ml notebooks**

**Explore the data set**

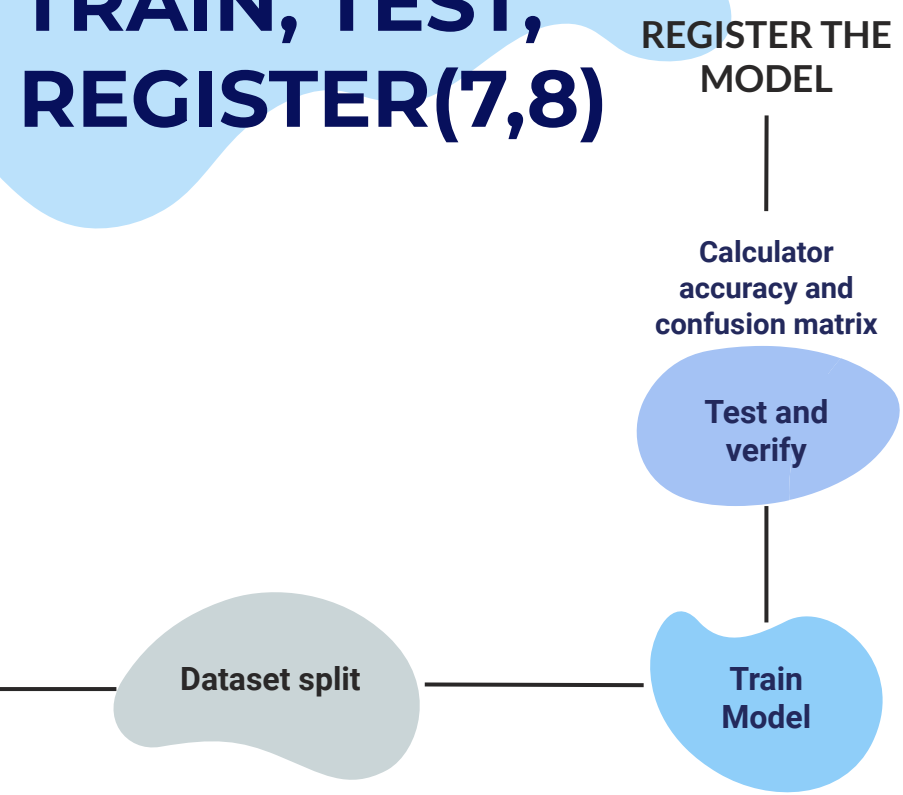
- **Features V1, V2, … V28 with PCA features**
- **'Time' and 'Amount' and 'Class' with non-PCA features**

**Normalize the data**

**reduce the redundancy**

*(kaggle.com, 2018)*

# TRAIN, TEST, REGISTER(7,8)




```
x_df = dataset.drop('Class', axis = 1).values
#remove the Class coloumn
y_df = dataset['Class'].values #use train split module to train and test using same database
X_train, X_test, y_train, y_test = train_test_split(x_df, y_df,test_size=0.2,random_state=66)

knn = KNeighborsClassifier(n_neighbors = 5)
model = knn.fit(X_train, y_train) #use training data to evaluate a model
knn_y = knn.predict(X_test) #knn_y would be the system evaluation based on training that it had
print('Accuracy score of the KNN model is {}'.format(accuracy_score(y_test, knn_y)))

joblib.dump(model, 'sklearn_knn_model.pkl')
ws = Workspace.from_config()
model = Model.register(workspace=ws,
model_name='my-sklearn-model',                # Name of the registered model in your workspace.
model_path='./sklearn_knn_model.pkl',  # Local file to upload and register as a model.
model_framework=Model.Framework.SCIKITLEARN,  # Framework used to create the model.
model_framework_version=sklearn.__version__,  # Version of scikit-learn used to create the model.
resource_configuration=ResourceConfiguration(cpu=2, memory_in_gb=4),
description='knn model to predict fraud trasnactions',
tags=None)
```

```
Accuracy score of the KNN model is 0.9994733330992591
['sklearn_knn_model.pkl']
```

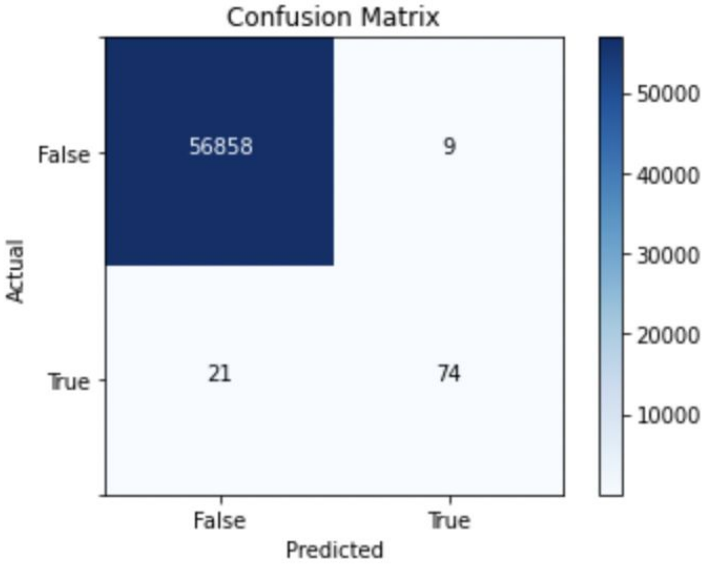
By using a train and split module, test data is tested against the same data it was being trained with

- K nearest neighbor algorithm
- Use 80% of whole dataset for model training

# CONFUSION MATRIX (8,9)

The confusion matrix shows cases instances where:

- Both the predicted and actual values were “1” (known as *true positives*)
- Both the predicted and the actual values were “0” (*true negatives*)
- The predicted and actual values differ (*false positives* and *false negatives*).




Colour intensity has direct relationship with number of correctly identified cases. This feature can help you identify a model that predicts accurately for all classes by looking for a diagonal line of intensely colored cells from the top left to the bottom right

*Sklearn.metrics.confusionmatrixdisplay*. scikit. (n.d.). Retrieved July 24, 2022

# BUSINESS CASE

## Payment Structure:

Subscription Based
$18/month

## Potential Market:

- Ever increasing as people become more and more aware of economic and internet safety risks
- Sales for subscription based safety products such as VPN's are rising (10)

## Microsoft Cost Calculator Estimation:

Logic App + API Management = $821/month

# BUSINESS CASE: COMPETITION

- The biggest credit card companies have begun to use AI to help combat Fraud (11, Bankrate.com)
  - American Express
  - Mastercard
  - Visa
- What do we do better?
  - Transparency
  - Alerts
  - Data Storage
  - Cost- Effectiveness with Azure

# POSITIVE IMPACT ON SOCIETY

Guarantees **stable finance** after **retirement**

**Boosts** employee **motivation**

Causes **economic growth**

Forms **credible** social atmosphere

Obviates **secondary victimisation**

**Increases** overall **technical skills** in financial protection in companies

# IMPACT MEASUREMENTS

Collecting **user feedback** through monthly optional poll

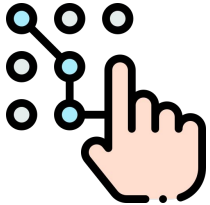
**Comparing the current statistics** of seniors fraud damage rate with the **past**

**Tracking** the usage of the AI tool and counting its **interactions** with users

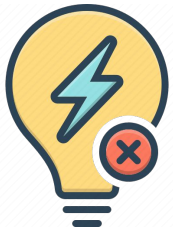
# NEGATIVE IMPACTS

Data is unbalanced

The system detects some bad patterns



Power Outage



financial expose & loss

Loss of Access to servers

Have downtime

Hacking

# HOW TO MITIGATE NEGATIVE IMPACTS CONT'D

**Availability Zone**

Decrease latency and **protect resources Failover outage** events (12)

**Azure portal**

**Resilient datacenter failures** and avoid network slowdowns

**Azure Monitor**

Set up **alerts** for key events that are related to users' specific resources (outages)

(Docs.microsoft.com/availability-zones, 2022)

# ETHICAL CONSIDERATIONS

## Transparency

- Be able to see stored data and the use relationship
- **Azure Machine Learning Workspace** (13)

## Privacy and Security

- **Protect** personal data
- Automated security email
- Needs **authentication** Method

## Reliability and Safety

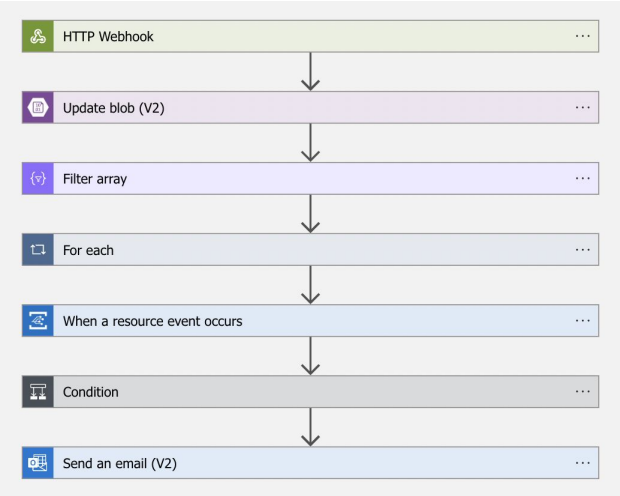
- **Regular** Monitoring and **inspection**
- **Stable** cloud **environment**

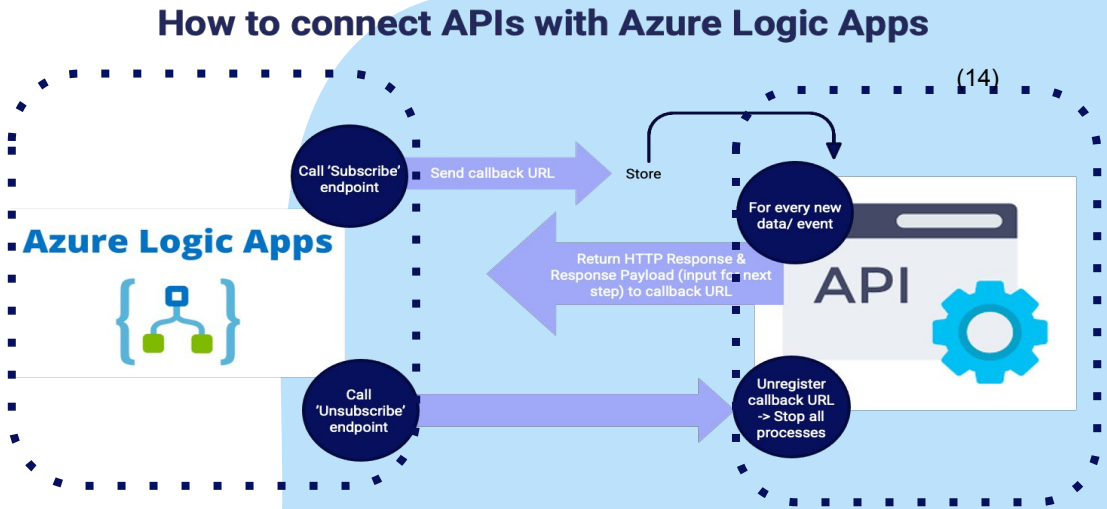
(Docs.microsoft.com/cloud-adoption-framework/innovate/best-practices/trusted-ai , 2022)

# FUTURE RESEARCH

Integration of Azure Logic App with custom API :

- Use the webhook trigger to request data from the API
- Use Web-endpoint to minimize the knowledge of any technicality




(Docs.microsoft.com/zure/logic-apps, 2022)

# THANK YOU FOR LISTENING!

# Bibliography


1. Federal Trade Commission. (2022, February). *Consumer Sentinel Network Data Book 2020*. Retrieved July 16, 2022, from https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-2020/csn_annual_data_book_2020.pdf
2. *Credit Card Fraud Statistics: What are the odds?* Dataprot. (n.d.). Retrieved July, 2022, from https://dataprot.net/statistics/credit-card-fraud-statistics/
3. Nilson Report. (2021, December). *Credit Card Fraud Worldwide*. Retrieved July, 2022, from https://nilsonreport.com/upload/content_promo/NilsonReport_Issue1209.pdf
4. IntSights Cyber Intelligence. (2018, July 11). New Research from IntSights Cyber Intelligence Finds 135% Year-Over-Year Increase in Bank Data for Sale on Dark Web Black Markets. Retrieved July 0, 2022, from https://intsights.com/press-releases/new-research-from-intsights-cyber-intelligence-finds-135-year-over-year-increase-in-bank-data-for-sale-on-dark-web-black-markets.
5. *Senior scam statistics 2022: Is elder fraud on the rise?* Comparitech. (2021, October 23). Retrieved July 26, 2022, from https://www.comparitech.com/identity-theft-protection/senior-scam-statistics/
6. ULB, M. L. G.-. (2018, March 23). *Credit Card Fraud Detection*. Kaggle. Retrieved July 26, 2022, from https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
7. joparga3. (2016, December 17). *In depth skewed data classif. (93% recall ACC now)*. Kaggle. Retrieved July 26, 2022, from https://www.kaggle.com/code/joparga3/in-depth-skewed-data-classif-93-recall-acc-now
8. Ebner, J., S, P., Z, K., & Pal. (2022, May 16). *How to use Sklearn train_test_split in python*. Sharp Sight. Retrieved July 24, 2022, from https://www.sharpsightlabs.com/blog/scikit-train_test_split/
9. *Sklearn.metrics.confusionmatrixdisplay*. scikit. (n.d.). Retrieved July 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.ConfusionMatrixDisplay.html

# Bibliography

10. Vigderman, A. (2022, July 22). *2022 research into VPN usage (updated statistics)*. Security.org. Retrieved July 24, 2022, from https://www.security.org/vpn/statistics/

11. Maxwell, T. (2022, June 15). *How major credit card networks protect customers against fraud*. Bankrate. Retrieved July 0, 2022, from https://www.bankrate.com/finance/credit-cards/major-credit-card-networks-protect-against-fraud/

12. Awysza. (n.d.). *Azure regions and availability zones*. Azure regions and availability zones | Microsoft Docs. Retrieved July 26, 2022, from https://docs.microsoft.com/en-us/azure/availability-zones/az-overview

13 . msteller-Ai. (n.d.). *Responsible and trusted AI - Cloud Adoption Framework*. Cloud Adoption Framework | Microsoft Docs. Retrieved July 26, 2022, from https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/best-practices/trusted-ai

14. Ecfan. (n.d.). *Create web apis & REST apis for azure logic apps - azure logic apps*. Create web APIs & REST APIs for Azure Logic Apps - Azure Logic Apps | Microsoft Docs. Retrieved July 26, 2022, from https://docs.microsoft.com/en-us/azure/logic-apps/logic-apps-create-api-app