**Transaction Categorization**

* **Topic**:

Transaction categorization is the ability to recognize the purpose of a transaction based on its description. For long, this process was done manually but now technology can do it efficiently.

This project will focus on Natural Language Processing using the power of Machine Learning to predict which category a transaction will fall into, given the description of the transaction.

* **Background & Business Problem:**

It is no surprise that financial companies need to help their customers organize their finances. Customers want to know what they spend their money on to keep balances in check. By categorizing transactions and building better customer engagement tools, Banks can help customers identify frequent purchases and subscriptions, sort income and activity liability with higher accuracy, and reduce credit risks.

This project will focus on Natural Language Processing using the power of Machine Learning to predict which category a transaction will fall into, given the description of the transaction.

**Goal** - Build a model to predict transaction categories using the 10 (ten) distinct categories that a transaction may fall into. The categories are as follows:

1. Communication Services
2. Education
3. Entertainment
4. Finance
5. Health and Community Service
6. Property and Business Services
7. Retail Trade
8. Services to Transport
9. Trade, Professional and Personal Services
10. Travel

* **Datasets**:

**Training Dataset** – The Training Dataset contains 40,000 unique transactions and their corresponding transaction categories. This data set is used to train the model.

**Test Dataset** – The Test Dataset contains 10,000 unique transactions with the transaction categories omitted. This dataset will be used to test the solution and identify the correct transaction categories.

* **Methods**:

To build this model, CRISP-DM methodology will be used and followed each stage from it to make sure right product is built with minimal issues.

* Business Understanding - Business objective will be determined by discussing with stakeholders and respective knowledge about domain will be gained to learn more about dataset and its sources.
* Data Understanding – Relevant data will be collected, and exploratory data analysis will be performed to gain more understanding from the data.
* Data Prep – In this stage data will be cleaned and transformed into format that ML model would accept it.
* Modeling – Model would be finalized based on problem statement and model will be trained and validated with training data.
* Evaluation – Trained model will be evaluated based on different metrics and result will be reviewed and if needed dataset will be changed and model will be re-trained with new changed datasets.
* Deployment – The final model will be deployed into production system, where it would help to predict right category from the given transaction.

This problem statement falls under classification techniques so different Machine learning and Deep Learning Classification Algorithms will be used to build this model.

* **Techniques** – The model would be built using Predictive Machine Learning Algorithms for classification and Deep learning techniques will also be used to build the model.
* **Ethical Considerations:**

The transactions data is captured and stored with real time and real customers so to protect customer PII information, the dataset has been scrubbed to replaces all transactions number and credit card numbers with 1’s.

* **Challenges:**

To use deep learning to train the model GPU enabled machines will be needed to train model fast and would save time but without GPU, it would be challenging to train model.

* **References**:
* <https://www.mindsumo.com/contests/a13fa0e9-517d-41be-8f8b-26f735038fb7>