Ivory (Xiangzi) Chen

647-989-3419 | ivory.chen.job@gmail.com | Linkedin | Github

OBJECTIVE

Data Scientist with 5 years of experience in data analysis and machine learning seeking a challenging role to utilize expertise in **Traffic Route** selection, **Banking Transaction** analysis. and **Customer Segmentation** and profiling

EDUCATION

University of Toronto

Master of Engineering, Mechanical and Industrial Engineering (**GPA:4.0**) Sep 2021 - Nov 2022

ML & Analytic Courses: Cloud-Based Data Analytics; AI in Finance; Algorithms & Data Structures; Data Analytics &ML; Deep Learning; MEng Research

Honours Bachelor of Science with High Distinction (Dean's List for 2020, 2021) Sep 2017 - Jun 2021

Specialty: Applied Statistics; **Minor**: Mathematics, Economics; **Courses**: data analysis; Statistical Consultation, Methods for Multivariate Data

CORE COMPETENCY

- ✓ **Data Analysis**: 5 years of experience in conducting exploratory, statistical, and **predictive** analysis using various techniques such as Regression, Deep Learning, NLP, Image Detection, and Hypothesis Testing.
- ✓ **Machine Learning**: Extensive experience in Modelling & Engineering with
  - Python packages such as Sklearn, Pandas, PyTorch, TensorFlow, GeoPandas, Numpy, Matplotlib, Seaborn, Shapely
  - Advanced SQL on PostgresSQL, ADF, Azure SQL DB, and Snowflake. Proficient in **R** and **SAS** for data analysis
  - Professional Visualization Design - **Power BI Dashboards**
- ✓ **Big Data**: Proven large datasets processing in Spark using PySpark and Spark-SQL in Databricks Cluster Computing.
- ✓ **Certification**: Certified in **Microsoft Azure Machine Learning** – Data Science Solution

WORK EXPERIENCE

Research Analyst | **Centre for Maintenance Optimization and Reliability Engineering,** Toronto, Canada Jan 2022 - Sep 2022

- Developed and implemented a **Best Safe Route Selection Model** for hazardous material transportation utilizing an extended SARSA reinforcement learning to determine the dynamic best route for trucks, resulting in a 30% decrease in overall risk
- Preprocessed four months of traffic, weather, and population data of GTA using Databricks Cluster Computing; Transformed raw data into evidence-based insights and ensured data accuracy/completeness by performing data quality checks and conducting data cleaning before ingestion; Pipeline design of getting accurate geospatial data for successful implementation
- Led **road network** analysis and proposed a **risk assessment** approach for optimal transportation routes. Facilitated timely project completion and fostered team collaboration by delivering weekly progress reports during SIT runs.

Data Analyst, Intern | **Chuang Chuang Culture Communication Co., Ltd.** Shanghai, China May 2019 - Aug 2019

- Developed customer profiles by analyzing **demographics** and **purchasing behavior** using R Studio. Insights were provided to stakeholders to design a new consulting product which generated over 0.8 million US dollars since its launch in 2019.
- Collaborated with the marketing team to create potential client lists based on distinct customer feature clusters identified using K-Means clustering. Generated on average 30% improvement in **email click-through rates**.
- Increased customer satisfaction score by 15% by developing a customer satisfaction dashboard with Power BI for management.

PROJECTS

CRM Transaction Analysis – Retail Banking Credit Card Promotion, Github Apr 2023 - May 2023

- Developed a DataModeler class in python for bank transaction analysis, including feature engineering, fitting and testing to targeted business categories for retail banking credit card promotion through clustering and quantitative analysis,
- Delivered a Customer Segmentation Dashboard in Power BI, improving campaign targeting efficiency by 30%.

Workforce Data Extraction with Azure Cloud Platform, Github Mar 2022 - Apr 2022

- Implemented an efficient data ingestion process with metadata driven pipelines in Azure Data Factory to transfer csv files from Azure Blob storage to Azure SQL database; established scheduled triggers to automate the process.
- Utilized T-SQL queries in Azure SQL Database to analyze Fact data by occupation, gender & time Dimensions

Vehicle License Plate Detection, Github Jan 2022 - Feb 2022

- Utilized Keras to design a VGG16 Convolutional Neural Network model for license plate detection. Achieved an impressive 88.1% accuracy rate in localizing bounding boxes of license plates in images.
- Built and fine-tuned a CNN classification model to identify the letters and digits on license plates with a 76.6% accuracy rate.
- Used computer vision library to create visualizations of the bounding box and character recognition results. Utilized image and video demonstrations to showcase the recognition accuracy from multiple angles and improve overall accuracy by 8%.

Canadian Election NLP Sentiment Classification Analysis, Github Sep 2021 - Dec 2021

- Utilized NLP techniques to analyze Twitter data during the 2021 Canadian Election, tested various algorithms including Logistic Regression, Decision Tree, Random Forest, KNN, SVM, Naive Bayes, and XGBoost to classify tweet sentiment. Achieved 97.4% accuracy in predicting public attitudes toward the election by implementing the best-performing model.