

PIZON SHETU

pizon.skhan@gmail.com
(917) 340-3347

www.linkedin.com/in/pizon-shetu/
www.github.com/izinex/

PROFESSIONAL EXPERIENCE

Data Scientist – Webster Bank (June, 2022 – Present Stamford, CT):

- Led the extraction of historical credit approval data to develop in-house **PD** and **LGD** models, saving \$1.5M annually by replacing Moody's models.
- Developed a Python-based algorithm and executable software to detect and count bias words in appraisal documents, reducing legal and reputational risks for the commercial line of business.
- Using **SQL**, **SAS** and **Python** perform end-to-end development, back-testing, and documentation of models across diverse portfolios, orchestrating internal and external data consolidation by filling missing data using 3rd-party data (TREPP) via nearest neighbor clustering (**KNN**).
- Aligned driver variables with default and loss patterns, ensuring effective representation of the bank's risk landscape in Commercial Real Estate (CRE), Sponsor & Specialty (S&S), and Asset Based Lending (ABL) models. Collaborated closely with IT, data teams, and credit underwriters to seamlessly communicate and align key data elements with business needs, ensuring accurate reflection of product consumer requests in the final product.
- PD and LGD models developed on **Binned Variables Logistic Regression using Weights of Evidence**; Models performed with an ROC of **~89%** and **~70%** for PD and LGD respectively for CRE.
- Conduct stress tests (model performance), validating predictor variables' relevance with new data, and code reviews.

Data Science Mentor – Springboard (2023 – Present, New York, NY):

- Mentor Springboard students by providing coding support, goal-setting guidance, and career advice, enhancing their data science skills and career readiness. Cover a wide range of topics including data wrangling, data visualization, exploratory data analysis, statistical inference, and machine learning.

Data Scientist – Whiterock.ai (Jan, 2022 – May, 2022 Manhattan, NY):

- Conducted exploratory data analysis (EDA) and performed Extract, Transform, Load (ETL) on large real-estate datasets to identify key insights and features on past and present markets, to feature engineer key risk drivers for home prices.
- Automated data ingestion from BlackKnight and other sources using Apache Airflow and GCP, reducing processing time by 20% and improving data quality.

Junior Data Scientist – ProMarketingHub (July 2020 – Oct 2021 Queens, NY):

- Performed ETL processes and customer segmentation using **k-Means clustering** in Python, enhancing targeted marketing strategies for a startup marketing firm.

EDUCATION

Springboard Data Science Bootcamp – Online

Completed a comprehensive program in the full Python Data Science Stack, including Data Wrangling, Statistical Inference, Supervised and Unsupervised Machine Learning, Deep Learning, SQL, A/B Testing, etc.

Queens College Bachelor's in Computer Science and Applied Mathematics - NY, Queens

Relevant Coursework: Object-Oriented Programming, Data Structures and Algorithms, Database Systems, Computer Architecture, Software Engineering, Internet/Web Technologies, Theory of Computation, Probability and Statistics, Bayesian Modeling, Linear Algebra, Linear Programming, Advanced Calculus, Machine Learning in R, Blockchain Mathematics.

PROJECTS

Convolutional Neural Network for Image Recognition – Classification

- Developed and implemented a CNN using Keras API for bird species classification, achieving 94% accuracy, and improved to 98% through Transfer-Learning with VGG16 and hyper-parameter tuning.

New York Housing Price Prediction – XGBoost Decision Tree

- Cleaned 75K invalid and missing data points in Zillow's housing data using advanced imputation techniques (MICE), improving NYC housing market accuracy. Assessed predictive models (Linear Regression, RandomForest, KNN) using Mean Absolute Error. Optimized XGBoost achieved the highest accuracy through hyper-parameter tuning.

TECHNICAL SKILLS

Languages: Python, SQL, SAS, R, Java, C++, Excel, LaTeX

Technologies/Frameworks: Git, ABBYY, Databricks, Google Cloud Platform, Scikit-Learn, OpenCV, Pandas

Soft Skills: Strong Communication, Cross-functional Collaboration, Project Management, Mentoring