

PIZON SHETU

209-22 Whitehall Ter
Jamaica, NY 11427
(917) 340-3347

pizon.skhan@gmail.com
www.linkedin.com/in/pizon-shetu/
www.github.com/izinx/

PROFESSIONAL EXPERIENCE

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

- Utilized quantitative analysis to implement credit risk management models under CECL guidelines. Analyzed diverse financial data and applied advanced techniques to extract valuable insights, contributing to the decisions made by models.
- Built efficient data processing pipelines for parsing and cleaning large population of documents using regex and NLP, enabling data-driven selection of optimal sample populations for extraction.
- Utilized OCR technologies (Pyasdxatesseract), NLP, and ABBYY software to process and extract 36 data elements from over 1 million Word and PDF documents for model development.
- Contributed to the development of the Dual Risk Rating Model by engineering predictor and response variables, leading to the creation of accurate and efficient predictive models.
- Played a key role in team management and leadership, leading the project from proof-of-concept to development and taking ownership of the project's success.

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, resulting in significant improvements in the accuracy of predictive models.
- Automated incoming data from various sources using Apache Airflow, GCP BigQuery, and Google Cloud Storage, reducing processing time by 40% and improving data quality.
- Collaborated closely with the CTO to maintain daily model clusters, build predictive models for real-estate products, and research different models such as Neural Networks, demonstrating expertise in data science and machine learning.

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

- Managed and stored user data on cloud database, and performed daily ETL, data cleaning, and preprocessing.
- Worked closely with Sr. and Lead Data Scientists to generate and test hypotheses related to product engagement.
- Defined real-time customer data needs, evaluated data quality, and determined suitability for use.

Data Analyst – Centerplate (2016-2020 Elmont, NY):

- Drove customer engagement by conducting ETL on a large dataset of over 10K clients, uncovering critical insights and trends to inform decision-making and drive growth.
- Boosted customer orders by 13% through the strategic implementation of coupons and combo deals, resulting in increased revenue and profits.
- Optimized pricing strategies for food and beverage offerings based on sales data, leading to an increase in revenue.
- Collaborated with stakeholders to provide compelling data visualizations highlighting customer habits and key findings, enabling informed decision-making, and driving business growth.

EDUCATION

Springboard Data Science Bootcamp – Online (2021):

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 (2015 – 2020):

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

- Designed and built a complex neural network leveraging the Keras API for classification, achieving an impressive 94% accuracy in identifying 315 species of birds.
- Elevated the model's performance through the application of Transfer-Learning with VGG16 and meticulous hyper-parameter tuning, leading to an additional 4% increase in accuracy, culminating in a remarkable 98% prediction accuracy.

New York Housing Price Prediction – XGBoost Decision Tree

- Implemented advanced imputation techniques such as MICE to clean over 75K invalid and missing data points in Zillow's housing data, leading to a more accurate representation of the NYC housing market.
- Demonstrated proficiency in evaluating multiple predictive models including Linear Regression, RandomForest, and KNN, utilizing Mean Absolute Error as the evaluation metric. Conducted extensive hyper-parameter tuning on each model to optimize performance, ultimately achieving the most accurate results with XGBoost Gradient Boosting.

TECHNICAL SKILLS

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

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