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/izinex/

PROFESSIONAL EXPERIENCE

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

- Implement credit risk models under CECL guidelines. Analyze diverse financial data and apply advanced techniques to extract valuable insights, contributing to the decisions made by models.
- Lead a team comprised of both consultants and internal members, overseeing the extraction of data from historical documents. Successfully guide the project from the initial proof-of-concept stage to full-scale development, assuming complete ownership of the project's overall success
- Facilitate effective communication between cross-functional teams of model builders and Credit Risk subject matter experts (SMEs) to collaboratively develop predictor and response variables for the Dual Risk Rating Model.
- Built efficient data processing pipelines for parsing and cleaning large population of documents using regex and NLP.
- Utilize OCR technologies (tesseract), ABBYY software, and Python (Tabula/Camelot) to process and extract data from tables and contextual text in unstructured documents.

Data Science Mentor – Springboard (2023 – Present, Remote):

- Assist Springboard students in their Data Science journey, providing coding support, goal setting guidance, and career advice during weekly video calls. 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 incoming data from BlackKnight and other various sources using Apache Airflow and Google Cloud Platform (GCP) reducing processing time by 20% and improving data quality.

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

- Defined customer needs using data-driven methods for a startup marketing firm. Evaluated data quality and determined suitability. Segmented customers using k-Means clustering, allowing businesses to market to their target audience.

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

- Work with stakeholders and management to boost sales revenue by 13% through the strategic implementation of coupons and combo deals. Provide analysis and reports on weekly sales, inventory and sold products.

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

- Developed and implemented a neural network using the Keras API for bird species classification, achieving an impressive 94% accuracy in identifying 315 species. Further enhanced the model's performance through Transfer-Learning with VGG16, and hyper-parameter tuning, resulting in a 98% prediction accuracy.

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

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