PIZON SHETU

www.linkedin.com/in/pizon-shetu/ pizon.skhan@gmail.com (917) 340-3347

PROFESSIONAL EXPERIENCE

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

- Spearheaded extraction of crucial historical credit approval data for in-house PD and LGD models. Conducted research, integrating relevant 3rd-party data via nearest neighbor clustering, addressing missing data. Resulted in $1.5M annual savings by transitioning from costly Moody's models.
- Orchestrated internal and external data consolidation for model development. Collaborated with Credit Risk and Loan underwriters to align key data elements with business needs. Filled data gaps using Trepp loans, bridging credit expertise and technical implementation.
- End-to-end development, back-testing, and documentation of PD and LGD models Commercial Real Estate (CRE). Aligned driver variables with default and loss patterns, ensuring the models effectively represented the bank's risk landscape. Worked closely with business units to validate the alignment of models with actual business scenarios, culminating in the successful creation of PD model with ROC of 89% and LGD with ROC of 68% using Logistic Regression.
- Develop Sponsor & Specialty (S&S) PD and LGD model with ROC of 71% and 63% respectively using Logistic Regression; building 8 different scorecards for each different industry in the S&S portfolio by standardizing model metrics.
- Initiated the transition of legacy CECL models from SAS to Python, managing code conversion and exploring automation solutions with airflow. Conducted stress tests on historical models, validating predictor variables' relevance with new data through a weight of evidence approach.

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

- 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 (July 2020- July 2021 Queens, NY):

- Utilized data-driven methods to define customer needs for a startup marketing firm, employing k-Means clustering to segment customers to enhance targeted marketing.
- ETL process on multiple datasets in Python to compile customer data to supplement unsupervised k-Means clustering models

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, LaTeX

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

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