PIZON SHETU

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

PROFESSIONAL SUMMARY

Dedicated and motivated learner with experience working with large datasets, extracting meaningful insights and build intuitive models to help with decision making, allowing businesses to grow with their objectives in mind using data-driven solutions.

EDUCATION

Springboard Data Science Bootcamp – Online (2021):

Performed and implemented full Python Data Science Stack, Data Wrangling, Statistical Inference, Supervised and Unsupervised Machine Learning, Deep Learning, SQL, A/B Testing, etc.

Queens College: Double Major - NY, Queens (2015 – 2020):

- **Bachelor of Arts in Computer Science**

  Relevant Coursework: Object-Oriented Programming, Data Structures and Algorithms, Database Systems, Computer Architecture, Software Engineering, Internet/Web Technologies, Theory of Computation.

- **Bachelor of Arts in Applied Mathematics**

  Relevant Coursework: Probability and Statistic, Bayesian Modeling, Linear Algebra, Linear Programming, Advanced Calculus, Machine Learning in R, Blockchain Mathematics.

PROFESSIONAL EXPERIENCE

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

- Perform Quantitative analysis on various financial data to facilitate model implementation in credit risk management.
- Use machine learning to implement OCR text mining and image processing to extract data for model building.
- Develop and maintain models under strict CECL guidelines and regulations.
- Extract/Scrape unstructured data from over 1 million PDF files to structured CSV tables for model development by training ABBYY software using OCR's image recognition and NLP.
- Feature engineer predictor variables and response variables to build predictive credit approval/risk models from semi, and unstructured data set.
- Collaborate and communicate with other teams and executives to solve problems and research new solutions.

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

- Perform ETL, EDA on expansive real-estate datasets to identify key features and insights on past and present markets.
- Automate new incoming data from different vendors and incorporate to existing database and cloud storage using Apache airflow, GCP BigQuery, and Google Cloud Storage.
- Coordinate closely with CTO to upkeep daily model clusters, build predictive models for different real-estate products such as price and rent prediction, and research different models such as Neural Networks.

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

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

RELEVANT PROJECT EXPERIENCE

Designed and built Convolutional Neural Network for Image Recognition – Classification

- Utilized Keras API to build a complex neural network which can classify 315 different species of birds.
- Leveraged Transfer-Learning with VGG16, it was able to classify with a 94% accuracy.
- Hyper-parameter tuning the model through Bayesian Optimization and freezing/unfreezing CNN layers led to a further gain of 4% and result in 98% accuracy in predictions.

New York Housing Price Prediction – XGBoost Decision Tree

- Cleaned and implemented MICE imputations on over 75K invalid and missing data for Zillow's Housing data
- Provided key insights and underlying distribution into NYC housing market, such as wealth gap, cost of homes, and even migration of native residents
- Compare and build different predictive models such as Linear Regression, RandomForest, KNN, and hyper-parameter tune each model to find best result using MAE as a metric, Gradient Boosting yielded the best results.

TECH STACK

Python: Pandas, NumPy, OOP, SciPy, Statsmodel, Tensorflow, Keras, Flask, requests, PyTorch

Machine Learning: Neural Networks, Regression, Decision Trees, RandomForest, Classification, RandomForest, NLP, Git Scikit-Learn, OpenCV, R, ABBYY.