Song Chen (Chester) 3308 24th St, Vernon, BC, Canada | 604-889-1206 | ycbq999@hotmail.com

Summary

- Senior Data Analyst with 5+ year experience in ETL, data pipelining, and data visualization. Proficient in translating large raw datasets into actionable insights for Power BI applications.
- Technical Skills encompass big data processing, Azure Synapse Notebook, SQL, and web scraping. Well-versed in efficient data extraction and analysis.
- Passionate about MLOps and Deep Learning. Dedicated to integrating machine learning models and developing end-to-end applications for time series data.

Technical Skills

Programming language: Python, R, C++, SQL, VBA

Data Science:

- **Data Visualization:** Power BI, Matplotlib, Seaborn, Plotly, Tableau, ggplot2
- **Machine Learning:** Time Series Machine Learning, Deep learning with Pytorch, Scikit-Learn, Customer segmentation with K-Means clustering, Knowledge of common supervised machine learning models (SVM, Logistic Regression, Random Forest, XGboost, etc)
- **Big Data Analysis:** hands on PySpark data processing, knowledge of MLlib.
- **Data analysis**: Advanced Excel, VBA, Web Scrapping with Beautiful Soup, Selenium

Cloud Computing: Hands on Azure applications, Fabric, Data Factory, Synapse, App Service. Knowledge of AWS and GCP

App Development: MLOPs, MLflow**,** Flask, Shinny, Docker

Statistics: Hypothesis Test, Principal Component Analysis, Knowledge of A/B testing

Experience

PRICING ANALYST (SENIOR) |KALTIRE |OCT 2022- CURRENT

- Engineered data-driven solutions to dissect product cost, inventory, and pricing dynamics, providing a comprehensive view of the market landscape.
- Pioneered the development of automated Power BI sales reports, offering near real-time sales analytics and insights to the team.
- Conducted robust statistical analyses on diverse pricing scenarios, paving the way for the optimization of pricing strategies, increased up to 10% sales YvY
- Orchestrated daily data explorations on Azure Synapse Notebook using PySpark, streamlining ETL processes.
- Developed machine learning models including regression, clustering, XGBoost, deep Learning to predict sales trends and pricing strategies. Increased efficiency of pricing suggestions and improved accuracy by 20%.
- Presented pricing strategies and insights to senior management and stakeholders.

DATA ANALYST |PHASE ANALYZER COMPANY|JUN 2015 - OCT 2022

- Developed graphic data analysis with Excel VBA script and Python to study repeatability and reproducibility of equipment run tests. Implemented an automatic plotting feature within the Excel report, enhancing data visualization and aiding in the interpretation of complex data sets.
- Automated the web scraping process, enhancing efficiency and accuracy in collecting product pricing data across multiple webpages. Developed robust scripts in Python, leveraging Selenium for dynamic page interactions and Beautiful Soup for HTML parsing, ensuring seamless extraction of relevant price details.
- Implemented a K-Means clustering algorithm for customer segmentation, developed distinct service strategies for each segment, resulting in maximized service efficiency and increased revenue. Increased NPS score 20%
- Conducted ETL pipeline and statistical analysis to compute the chemical properties of petroleum products. Monitor system output and maintain the repeatability of the final output.

Education

MASTER|MAY 2022 |UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

- **Major: Computer Science, Data Science** (GPA 3.8)

BACHELOR|MAY 2014|UNIVERSITY OF BRITISH COLUMBIA

- **Major: Electric Engineering** (GPA 3.5)

Certificates

Azure Data Engineering (DP-203)

Projects

LSTM Deep Learning on Readmission prediction

- Spearheaded an independent deep learning project utilizing the MIMIC III database, focusing on a cohort of over 4000 patients diagnosed with Congestive Heart Failure (CHF).
- Developed and presented a cost-sensitive formulation of a Long Short-Term Memory (LSTM) neural network, incorporating expert features and contextual embedding of clinical concepts to enhance predictive accuracy.
- Conducted a comprehensive evaluation of each model element's contribution to prediction performance, employing metrics such as ROC-AUC and F1-measure.
- Concluded that the integrated model, incorporating all key elements, demonstrated superior discrimination ability with metrics including AUC: 0.91, F1: 0.53, and Recall: 0.89. This outperformance was observed across at least two evaluation metrics when compared to reduced models.
- https://colab.research.google.com/drive/1HYxEd3Li_PZBcB-32TkcSEGrgkvecw6Q?usp=sharing

MLOPs Object Detection App

- Designed and implemented a comprehensive facial gender detection application using Python, OpenCV, Flask, and Docker, with deployment on Azure App Service.
- Applied cutting-edge computer vision techniques, incorporating OpenCV and an SVM learning model to train on a dataset of over 4000 pictures. Achieved an impressive accuracy rate of over 82% in predicting the correct gender based on facial features.
- Developed a user-friendly application with Flask, enabling users to upload pictures. The application efficiently detects faces in the uploaded images and determines the gender for each detected face.
- Containerized the entire application using Docker, ensuring consistent deployment across various environments.
- Uploaded the Docker container to the Azure Image Registry. Successfully deployed the containerized application on the Azure App Service, achieving scalability and optimal resource utilization.
- ycbq999/Face_Recognition_MLOPs (github.com)

Movie Recommendation App

- Conceptualized, designed, and developed an interactive web application using the R language and Shiny package, leveraging the GitHub dataset sourced from Netflix.
- Conducted thorough exploratory data analysis, identifying genres and ratings as pivotal factors in attracting new viewers to the streaming platform.
- Successfully deployed the recommendation system with a user-based collaborative recommendation algorithm, allowing for personalized content suggestions based on user preferences.