Sunny Deng

Tel: 519-781-7023 **Email:** sunnydeng.work@gmail.com **LinkedIn:** https://www.linkedin.com/in/zixuansunnyd/

TECHNICAL SKILLS

- Programming languages: **Python, Java, JavaScript, Shell, SQL, R, Matlab, SAS**
- Database: MySQL, PostgreSQL, Redis, DynamoDB, Amazon Redshift
- ETL and Big Data: **AWS Glue, RedShift, EMR, Athena, Apache Airflow, Spark, Kafka**
- Machine Learning: KNN, SVM, Random Forest, Recommend System, CNN
- Python Packages/framework: Scikit-learn, Keras, Pandas, Numpy, Plotly, Pytorch
- Additional Software: **AWS RDS, EC2, S3, Lambda**, **Bitbucket**, **CICD** (Bitbucket + Git); **Jira, Confluence**

WORK EXPERIENCE

ARTICLE **Vancouver, BC**
Data Engineer *Jan 2022 – Aug 2022*

- Implemented, maintained multiple **PySpark data pipelines** up-time of **99.8%** while ingesting streaming and transactional data across **8 different primary data sources, 2 third party APIs** using various **AWS Service such as Redshift, S3, Glue** for **Tableau, Periscope, and Redash** reports.
- **Scheduled and designed ETL** workflows across **large datasets** using **Apache Airflow**, which reduced manual workload by **25% monthly**.
- **Improved** existing **DTS integration patterns** that allow import and pass data from new sources.
- Designed and created enterprise level data models from different dimensions to support business requirements.
- Collaborated with **data scientists to implement, maintain and adjust** end-to-end data pipeline to automate the **AWS Forecast workflow**, which saved planners **~8-10 hrs/week**.
- Migrated existing **30 reports and dashboards** so that they are compatible with the new database, that controls the influence of database migration to the lowest level.
- Communicated and collaborated with reports users and owners from finance, supply chain, category management and eCommerce to meet business requirements.
- Created readable technical documentation for **new integration pattern, Pyspark troubleshooting and Article's data platform architecture diagrams** for broader development team.

PROJECTS

FACE MASK DETECTION *May 2021 – Aug 2021*

- Trained **Convolutional Neural Network** with Keras.
- Utilized **YoloV4** and **Faster RCNN** for mask detection.
- Reached approximate **mAP (mean Average Precision) 98%.**

MINI KEYWORD SEARCH TOOL FOR LOCAL FILES *Jan 2021 – Apr 2021*

- Exhibits the functionality of page-rank algorithm on **20,000 HTML input files**.
- Implemented **Apache Spark** (mapper and reducer functions) to get inverted index.
- Realized result ranking using **TF-IDF cosine similarity** measure.
- Developed a simple interface for giving **keyword queries**, that could display matched results.

PREDICTION OF EXTREME WEATHER EVENTS *Sept 2019 – Apr 2020*

- Analyzed the pattern of weather dataset in R to predict future extreme weather events.
- Trained **LDA, QDA, Decision Tree and Random Forest** models.
- Evaluated models performance and reached approximately **98 % total accuracy and 75% precision**.

EDUCATION

UNIVERSITY OF VICTORIA **Victoria, BC**
M.Eng. Applied Data Science GPA: 8.33/9 *Jan 2021 - Aug 2022*
Coursework: Algorithms & Data Models, Systems for Massive Datasets, Optimization, Data Mining

WILFRID LAURIER UNIVERSITY **Waterloo, ON**
BA. Mathematics, Concentration in Statistics and Data Analytics *Sept 2016 - Aug 2020*
Coursework: Statistical learning, Regression Analysis, Data Structures, Mathematical Statistics, Time Series Analysis