Jerry Sun

https://github.com/jerrysun103 jerrysun222@gmail.com
https://linkedin.com/in/jerry-sun-890306127 +1-647-676-1833

Education

- **University of Toronto - St.George Campus** Toronto, Canada
  Honours Bachelor of Science in Computer Science and Statistics; cGPA: 3.8/4.0 *Sep 2016 – June 2021*

Programming Skills

- **Languages**: Python, Scala, C, SQL, R, Java, Matlab
- **Technologies**: Apache Spark, Git, MongoDB, PostgreSQL, Jupyter Notebook, Hadoop

Experience

- **Ontario Teachers' Pension Plan** Toronto, ON
  Quantitative Developer *Sep 2018 - Sep 2019*
  - **Risk Analytics Platform**: Developed an ETL pipeline for market risk analysis and active risk analysis based on Monte Carlo Simulation using Python for data wrangling and calculation, R for visualization, and SQL query and MDX query for data extraction. Automatically generated 1-year risk numbers for the 200 billion CAD net asset fund.
  - **Robotic Process Automation**: Utilized Python to create a Robotic Process Agent for a drill-down risk analysis on daily 30GB of unstructured data, which reduced total process time by 80%.
  - **Data Warehousing**: Set up the schema design and management of SQL database that are used by 9 team members.
- **Prmia Risk Management Challenge** Toronto, ON
  Canada Regional Final Round Participant *Spring 2019*
  - **Data Collection**: Collected 50-year Natural Gas price historical data and 50-year Electricity price historical data using Bloomberg Terminal. Also collected 50-year temperature data from government website.
  - **Forecasting**: Created an ARIMA time series model to fit temperature data and applied decision tree model for price forecasting.
  - **Scenario Simulation**: Designed and implemented scenario simulator for a power plant operation by using trained temperature model, Natural Gas pricing model, and Electricity pricing model. Also, conducted strategy for portfolio hedging and portfolio optimization.

Projects

- **Spark NLP Sentiment Project**
  Twitter Sentiment analysis on real-time streaming *Feb 2021*
  - Consumed data from and pushed data into MongoDB by leveraging MongoDB Spark Connector.
  - Utilized Spark NLP to predict sentiment polarity on 1.6 million tweets, achieving 83.6% accuracy.
  - Implemented Scala code to find the most popular real-time Twitter hashtags by using Apache Spark Streaming and Twitter developer API.
  - Classified tweets sentiment polarity on streaming data for any user-defined topic.
  - Generated the target tweets sentiment summary over a user-defined time window by processing the data in MongoDB.
- **Toronto Condo Market Analytics**
  A systematic way to observe Toronto Condo market *Dec 2020*
  - Developed an efficient Python-based web scraping workflow to collect all real estate data for Toronto Condo Market.
  - Visualized sale data and sold data using Plotly and Matplotlib to analyze the supply-demand trend over time which generated 15 market change plots for each run.
  - Trained, tested, and deployed an XGBoost model, a Random Forest model, and a KNN model for Condo pricing which resulted in a $40,000 Mean Absolute Error for all sold condos in Greater Toronto Area from December 2020 to March 2021.
  - Analyzed market price data to find active undervalued Condos which reduced the browsing time from a couple of hours to a couple of minutes.