JIAN YANG

yangjian106@hotmail.com | (587)-968-3398 | Calgary, AB, Canada

PROFILE

- **Analytical**: Data Science practitioner with 3+ years of extensive experience in various analytical roles. Hands-on experience in statistical analysis, predictive modeling, machine learning algorithms, and data visualization using SAS, R and Python.
- **Communication**: Effective communicator with the ability to simplify complex concepts and interpret data to both tech and non-tech audiences in a clear and concise manner.
- **Motivated:** Data savvy, passionate about extracting insights from real-world datasets by harnessing the power of data analytics, always ready to develop new knowledge/skills.

CORE COMPETENCE

Statistics: Probability Theory, Generalized Linear Models (GLM), Hypothesis Testing, Maximum likelihood estimation, Monte Carlo Simulation, Sampling Method
Data Analytical Tools: SAS, SQL, R, Python, Hadoop (basic)
Machine Learning Models: Logistic Regression, Random Forest, Gradient Boosting (XGBoost), K Means Clustering, Recommender System, Neural Network, Sentiment Analysis
Data Visualization: Tableau, Pivot Table (Excel), Matplotlib, Seaborn (Python), ggplot2 (R)

PROFESSIONAL EXPERIENCE

Data Scientist Jan, 2018- Recent
ATCO Gas, Load Research & Forecasting

- Managed the end-to-end process of project scoping, data collection and integration, model development and validation, statistical analysis within a cross-functional team.
- Investigated and assembled large utility data from multiple resources by developing and maintaining scripts and queries in SAS to import, clean, transform and merge data.
- Forecasted short-term (daily) gas demand for our 1.2 million customers using statistical models (Multiple Linear Regression) and other analytical techniques.
- Explored on alternative innovative Machine Learning-based models such as Artificial Neural Networks (ANN), Random Forest, XGboost to improve model performance.
- Detected the possible outliers and data anomalies of gas consumption data, investigated the root cause behind them and corrected them timely working with IT department.
- Created visually impactful dashboards using Excel (Pivot table) and Python (Matplotlib, seaborn) for daily forecast reports and month-end presentations.
- Wrote in-house queries in SAS to perform data wrangling, ad-hoc statistical analysis and automated the data manipulation and reporting process.
- Provided expertise on statistical concepts for the engineering, customer billing team and promoted a data driven culture across the ATCO Gas department.
- Interpreted and communicated quantitative findings (models, data) to fellow colleague, senior executives, and presented in public seminars and company-wide presentations.

Data Analyst/Statistician April, 2017-Dec, 2017
Alberta Health Services, Alberta Children's Hospital

- Designed analytical methodologies for a Pan-Canadian project on Childhood Nephrotic Syndrome, provided statistical modeling and data output for publication and dissertation.

- Collaborated with physicians and other biostatisticians to help build and maintain a nation-wide clinical database by cleansing, updating and manipulating the data using R and SQL.
- Determined the association between the relapse rate, steroid dose and other risk factors (patient age, gender, ethnicity) using Poisson regression model.
- Predicted pathological type using supervised machine learning (random forest) algorithm with Python (Scikit-Learn) and investigated the relative importance of the patient features.
- Provided descriptive statistics and data dashboards through charts/graphs using Excel and R; wrote reusable R codes for statistical analysis and data quality control.

Statistical Analyst
Breast Cancer Society of Canada Feb, 2016-Dec, 2016

- Conducted in-depth research (quantitative and qualitative) on breast cancer and analyzed the longitudinal data of 800 breast cancer patients in Tom Baker Cancer Centre.
- Utilized supervised machine learning techniques to design a model (logistic regression) to predict tumor type (malignant/benign) and analyze the the driving factors impacting the survival rate using principal component analysis with Python.
- Built innovative non-parametric models to predict gene expression level of two bio-markers which are important to the survival rate of breast cancer patients.

EDUCATION

Master of Science, Statistics
University of Calgary GPA: 3.72/4.00, top10% of class Sep, 2015-July, 2017

AWARDS AND CERTIFICATE

- Recipient of Entrance Scholarship for International Student ($48000)
- Transformative Internship Scholarship, University of Calgary ($8000)
- Certified Machine Learning (Deep Learning) Engineer (GPA: 97/100)
- Advanced SQL for Data Scientist Certificate by LinkedIn
- SAS Certified Base Programmer for SAS 9
- Tableau Essential Training Certificate by LinkedIn
- Contributed Speaker on Annual Meeting of Statistical Society of Canada

VOLUNTEER EXPERIENCE

External Officer, Alberta Artificial Intelligence Association Jan, 2018- Now
• Assisted in a series of service events to help members learn more about the AI theories and build their technical skills through boot camps and workshops.
Volunteer, Edmonton Humane Society Jan, 2018- Feb, 2019
• Volunteered with cat cuddling and dog walking, helped with shelter animal photography and helped with some fundraising events.

SKILLS

Computer skills: Proficient in Excel, Word, PowerPoint and Google G suite
Languages: English, Mandarin, R, Python
Hobbies: Tai Chi, jogging, Ping Pong, tennis, light hiking