Behnaz Jafari
E-mail: Behnaz.jafari@gmail.com Phone: 587- 968 6648
LinkedIn URL: www.linkedin.com/in/behnaz-jafari-ra Calgary AB, T3A 3H1

PROFESSIONAL SUMMARY

- Competent in healthcare data analysis, using datasets such as AHS administrative data and Kaggle open healthcare datasets, performing different statistical methods, survival analysis and applying machine learning algorithms on healthcare data using statistical software: R, Python, SAS and STATA,
- Competent in applying ML methods on multidimensional scientific (weather data NetCDF) data as well as data visualization using Google maps,
- Competent in data visualization used different methods: Heatmaps, Interactive maps, different graphs, plots and charts,
- Competent in systematic review and meta-analysis; used Covidence, EndNote and R;
- Experienced software developer, used VB, C, and database : SQL and Access
- Familiar with GitHub and Azure Databricks

EDUCATION

MSc in Biostatistics Sep 2017 – Jan 2020
University of Calgary, Calgary

Thesis: "Bias and Bias-Correction for Individual-Level Models of Infectious Disease"
Special foci: Statistics, Biostatistics, Data science
Courses (related to Statistics and Data Science): Health Data Science; Probability; Mathematical Statistics; Generalized Linear Models; Survival Analysis; Bayesian Statistics; and Computational Statistics.

MSc in Industrial Engineering
Khatam University, Tehran

Thesis: "Performance Measurement in Supply Chain Management"
Major: System Management and Productivity (GPA: 17.94/20)
Courses: Human Resource Management; Management Information Systems; Project Management; Supply Chain Management; Advanced Operations Research

BSc in Computer Engineering
University of Azad (Tehran Central Branch), Tehran

Major: Software Engineering (GPA: 15.47/20)
Major project: "Object-Oriented Designing and Programming with Personalities"
Courses: Assembly and Machine Languages; Data Structure and Algorithms; Numerical Computational Methods; Data Storage and Retrieval; Principles of Compilers Design; Principles of Database Design; Software Engineering

RELEVANT PROJECTS

Data Science Industry Fellowship for Albertans- Cybera Jul-Aug 2021
Used ML algorithms to predict localized demands to serve people in distress, using call centre data-211 in Alberta

- Used GitHub, Azure Databricks and Slack to collaborate with team members for the purpose of big data analysis.
- Cleaned , wrangled and visualized the data used spatiotemporal mapping for Calls, Referrals, Resources and Met and Unmet Needs datasets (830000 , 170000 , 67000 and 650000 records, respectively)- Used packages: Matplot, Geopanda, Folium, Zipfile with Python
- Used NLP to analyze text data used in follow up and referrals datasets- Used packages: Tensorflow, NLTK, Jason, Gensim and Sklearn with Python
- Determined and Predicted the amount of calls and resources during COVID crisis 2019-2021 using ML algorithms;

A Twitter-based flu surveillance system to predict a disease outbreak 2019 –2020

- Identified and tally flu-related tweets (20 MB tweets)
- Cleaned the tweets data using Scikit-Learn and NLTK (Natural Language Toolkit) libraries
- Developed a flu surveillance system using machine learning algorithms (Random Forest Classifier, SVM, KNN, Logistic Regression) with Python;

An infectious disease model in a Bayesian framework for simulating and predicting an infectious disease transmission 2018 –2019

- Simulated data from spatiotemporal individual-level infectious disease models,
- Estimated model parameters using Markov chain Monte Carlo technique in a Bayesian framework, used R to develop and implement the models,
- Data visualization using different tools such as heatmaps and charts;

A statistical model to predict German patients chronic heart disease 2017 –2018

- Cleaned and prepared German chronic heart disease patients' data (2500 patient's' records) for analysis using open health dataset Kaggle,
- Developed a logistic regression model to predict chronic heart disease; used STATA to implement;

WORK EXPERIENCE

Research Assistant & Data Analyst March 2020 —

Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary

Projects:

- Cleaning, linking and wrangling Canadian administrative health datasets using SAS for the purpose of predicting dementia risk index among Albertan adults (over 55) using ML methods (65 GB data consists of: Physician claims, Ambulatory / National ambulatory care reports, Hospital Discharges, Registry and Census, Pharmacy dispenses);
- Survival analysis, using Cox proportional hazard models, of administrative data for the purpose of developing dementia among elders more than 55 years old
- Systematic Review and Meta-Analysis of the Effects of Nature-Based Activities Health Outcomes of Older Adults; used Covidence, EndNote and used R for analysis,
- Meta-analysis of "Definitions and measurements of polypharmacy in observational studies using administrative databases: a systematic review and meta-analysis"; used R for meta-analysis,

Research Assistant & Data Scientist Mar—Nov 2020

Department of Mathematics and Statistics, University of Calgary, Calgary & TransAlta Corporation, Calgary

Project: Created a regime-switching vector autoregressive statistical model to improve short term (three to six hours) forecasting wind power of TransAlta company; used R language

- Collected 10 years hourly historical weather data of ERA5 dataset with 0.28' x 0.28' spatial resolution for 4 variables: wind (U and V speed), surface level pressure, geopotential height in NetCDF format (10GB data); prepared and cleaned the data for analysis
- Used PCA method for data reduction
- Used SOM algorithm and Davis-Bouldin index for data clustering; used K-mean for regrouping the clusters as the regimes
- Created a regime-switching vector autoregressive statistical model to forecast wind power
- Developed a forecasting aggregated wind power generation
- Developed a price quantity pairs model for the electricity price based on the level of wind power generation; Used MSE as an evaluation method for the model;

Software Engineer & System Analyst (R&D Group) 2006 — 2009

Iran Railway Company, Tehran

- Using VB and SQL, developed, monitored, maintained, and revised scheduled maintenance plan for tracks and bridges, which was successfully implemented and used in Track and Building Department,
- Designed a training system for employees that aligned with the objectives of each sub-department (e.g., reduce workplace injuries or train employees on a new machine). Constant evaluation of the effectiveness of the training system,
- Using SQL and C language programming, conceptualized, developed, and maintained database software for the company's entire computer hardware;

COMPUTER SKILLS

Statistical software:	R, Python, SAS, Stata
Computer programming languages:	C, VB
Databases:	SQL, Access

PROFESSIONAL DEVELOPMENT

- Taken online course "Meta-Analyses I" and "Meta-analyses II" from Statistics.com the Institute for Statistics Education 2021
- Taken online course "Working with Administrative Data" from the University of Victoria 2021
- Attended SAGES (SoTL Advancing Graduate Education in STEM) teaching and learning program
  Taylor Institute of Teaching and Learning, University of Calgary 2019