Dr. Sri Sripirakas

+1 (416) 474 7435 | sripirakas363@gmail.com | Github: sripirakas

Summary

A commercial data scientist with 5+ years of analytics experience and a well-awarded PhD in machine learning, with publications in top-ranked journals and conferences. Published my own algorithms in a popular open source tool, namely MassiveOnlineAnalysis (MOA) (Apache SAMOA). Demonstrated problem solving and creative thinking by effectively utilizing publicly available web data for fraud detection and finding new business opportunities at Tenzing. Delivered numerous end-to-end machine learning solutions with data engineering and visualization.

Achievements include creating pricing models that increased revenue by 15 million within 3 months and producing more accurate forecasts and insights for timely decisions that slashed operational costs by at least by 10% in manufacturing, planning and business processes at Fonterra.

Skills

Development	Machine Learning, Statistics, Data/Data Stream Mining, Data Engineering, NLP
Development	Data Structures and Algorithms, Object Oriented Concepts, Design Patterns
Programming	Python, R, Java, C#, C, C++, SAS, Javascript, BASH
Visualisation	R-Shiny, Tableau, QlikView, Microsoft PowerBI
Databases	SQL, NoSQL Databases (unstructured), PLSQL, Oracle, SQL Server, PostgreSQL, DynamoDB, MongoDB, HBase
Platforms	Spark, Cloudera, AWS, Google Cloud, Docker, Hadoop, Weka, MOA, Knime, Orange, Linux
Python Libraries	scikit-learn, sklearn, scipy, nltk, pandas, numpy, TensorFlow, keras, matplotlib, unittest
R Libraries	statmodels, dplyr, Random Forest, prophet, forecast, party, ggplot, plotly

Experience

Fonterra Group Limited, Auckland, New Zealand **06/2018 - Current**
Data Scientist

- Achieved more than $15 million net revenue within 3 months in a trial market by creating a price optimization model using regression models (Python and Tableau) for consumer products for Asia, Europe, Middle east, South America and Oceania regions
- Improved accuracy of short & long term forecasts of milk volume and composition by at least 20% at country and regional levels by developing time series models in R and Tableau
- Enabled reduction of millions in operational costs by identifying upcoming risks in manufacturing food products from creating an automated manufacturing insights awareness system that analyses millions of quality control data points and produces visualizations (R and Shiny) for all 200+ manufacturing plants
- Replaced an inaccurate-traditional demand forecasting method with an analytical model based on time series and regression (Python) using global dairy import & export and sales data for budgeting and planning purposes
- Rectified a failing random forest-based forecast model by treating co-variate shift in the underlying data set
- Improved accuracy by 15% and running time by 10% of the cost simulation models developed with Monte Carlo Simulation and Response-Survey Analysis in R
- Supervised peer-projects by validating modeling approaches and reviewing code and actively built an agile-based Data Science team culture
- Ensured expected business gains by identifying and translating fundamental drivers of business value into appropriate analytical problems, rather than solving problems at surface-level
- Formed awareness and trust on machine learning based-decision making process through Data Science community by doing workshops, demonstrations and presentations to staff

Tenzing Limited (formally KPMG Consulting), Auckland, New Zealand **01/2016 - 06/2018**
Senior Consultant - Data Science

- Enabled New Zealand's largest grocery retailer to customize promotions to increase sale revenue by creating an analytical tool to assess the effectiveness and the impact of promotions on affinity products using affinity analysis (association rule mining)(R and Shiny)
- Helped identify many online frauds/scams by developing a product (Python) to gather relevant financial news from social media, news channels, and other websites for a government financial market regulator (Web Scraping and text mining)
- Recommended methods to improve the effectiveness (Operational Analytics) of various machine learning models at a government tax department by finding underlying causes of ineffectiveness
- Completed Cloudera boot camp and participated in the development of in-house productionization platform using AWS, Docker, Spark, APIs and Hadoop

deFacto Software New Zealand Limited, Auckland, New Zealand **12/2015 - 01/2016**
Independent Contractor

- Commercialized a product based on my own algorithms namely SeqDrift1 and SeqDrift2 to detect abnormal signals from real time hardware logs of a server farm to automatically flag malfunctions and to predict them for preventive maintenance (Java and MOA-Weka)

Auckland University of Technology, Auckland, New Zealand **04/2015 - 04/2016**
Research Assistant

- Extended my PhD work to make it scalable for high dimensional and low memory requirements by modifying Discrete Fourier Transform of Decision Trees to use a divide and conquer-dynamic programming approach with Bayesian theory

Datamine Limited, Auckland, New Zealand **01/2015 - 04/2015**
Data Scientist - Intern

- Created a big data layer with Hadoop-MapReduce & HBase
- Developed web plug-ins with Django, Celery, Swagger and Ember.js for the analytics ecosystem

MillenniumIT | London Stock Exchange, Colombo, Sri Lanka **08/2010 - 06/2011**
Software Engineer

- Heavily applied design patterns, object oriented programming, high performance computing and highly scalable techniques
- Improved latency in data writing and designed fault-tolerant data storage

Education

Auckland University of Technology **08/2011 - 11/2014**
PhD in Computer Science

- Conducted extensive research on Real-time/Online Machine Learning in High Speed Data Streams
- Introduced four novel algorithms that are superior to state-of-art methods in terms of false positives, memory and run time complexity for detecting concept changes and to capture and reuse recurrent concepts in data streams
- Used various machine learning and statistical analysis techniques particularly concentration inequalities, sequential hypothesis testing, probability and distributions, discrete Fourier transform, classification and prediction with decision trees/Hoeffding trees and neural networks, Bayesian theory, sampling, data transformation and feature extraction methods etc. and implemented in C#
- Published the researched outcomes in Journal of Machine Learning, a top-ranked journal in ML and IJCNN, DAWAK, PAKDD conferences
- Published two change detection algorithms in Massive Online Analysis (MOA), an open source extension to Weka and received the Dean's Award for best doctoral research

University of Colombo School of Computing **06/2006 - 08/2010**
BSc in Computer Science, First Class Honors

- Constructed a statistical machine translation system for Sinhala and Tamil languages using Moses, Giza++ and SriLM toolkit with modifications to underlying algorithms to suit the languages

Publications and Presentations

JML - Journal of Machine Learning **2014**
Detecting concept change in dynamic data streams - **link**

IJCNN - The International Joint Conference on Neural Networks **2015**
Use of ensembles of Fourier spectra in capturing recurrent concepts in data streams - **link**

DAWAK - Data Warehousing and Knowledge discovery **2014**
Mining Recurrent Concepts in Data Streams Using the Discrete Fourier Transform - **link**

PAKDD - Pacific-Asia Conference on Knowledge Discovery and Data Mining **2013**
One Pass Concept Change Detection for Data Streams - **link**

ICTer - International Conference on Advances in ICT for Emerging Regions **2010**
Statistical machine translation of systems for Sinhala - Tamil - **link**

Honors & Awards

Dean's award for best doctoral research, PhD research **2016**
Laptop Scholarship and Summer Research Student Support, PhD research **2012**
Full Doctoral Scholarship and PhD publication stipend, PhD research **2011**