(416) 706-1805
chandana.karu@gmail.com

Chandana Karunaratne

chandana.ca

Toronto, ON M4J 3W6
linkedin.com/in/chandana-karu

PROFILE

Data analyst with MSc in data science and 7+ years of experience in applying statistical analysis to economic and market research. Proven track record of driving employer revenue growth and reducing employee labour costs via development of automated data analytics tools using Python and MySQL.

TECHNICAL SKILLS

- **Key Programming/Query Languages:** Python (Pandas, Numpy, Scikit-learn, GeoPy, others) and MySQL
- **Machine Learning:** natural language processing, classification, clustering, simulation modelling
- **Statistical Methods:** regression modelling, time series analysis, predictive modelling, significance testing
- **Data Visualisation:** static and dynamic visualisation models using Tableau, Matplotlib, Seaborn, Processing
- **Big Data Analytics:** Hadoop MapReduce, Hive, Spark SQL, PySpark, MongoDB, Cloudera

PROFESSIONAL EXPERIENCE

Lead Data Analyst – EvaSys (London / Edinburgh, United Kingdom) **Oct 2017 – Oct 2019**
EvaSys provides online survey software and data analytics to the higher education sector in the UK, mainland Europe, and North America (with headquarters in Lüneburg, Germany).

- Responsible for building data analytics tools, maintaining ETL (Data Extraction, Transformation and Loading) processes, and automating various tasks conducted manually.

- Responsible for developing predictive text classification tool using Python (Pandas, Scikit-learn) that categorizes survey feedback comments by thematic area based on machine learning models. The tool uses accuracy indicators from confusion matrices, including precision and recall scores, to compare the results from various classification models, including Multinomial Naive Bayes, Stochastic Gradient Descent, and Random Forest classifiers, among others.

- Developed sentiment analysis tool that visualizes degree to which survey feedback comments are deemed positive or negative while taking into account sentence structure and vocabulary of each comment using Seaborn and Pattern packages in Python.

- Developed and maintained ETL queries to i) extract structured and unstructured data from heterogenous sources like CSV files and Excel Workbooks, ii) transform data by cleansing, data mapping, and other processes, and iii) loading transformed data into in-house data warehouse using MySQL Workbench.

- Managed version control of Python and MySQL code using Bitbucket and Sourcetree.

Key achievements:

- Helped company generate over $2.2 million in annual revenue by leading in-house development of top-selling suite of text analytics tools, including text classification and visualization tools.

- Reduced process duration of multiple data analysis tasks that were manually conducted in Excel from 2 weeks to less than 1 minute by automating entire process in Python, helping company save over 1,200 man-hours annually among technical staff.

Technical Lead (Statistical) – International Labour Organization (Colombo, Sri Lanka) **July 2014 – June 2016**
The ILO is the United Nations agency specializing in labour-related issues and operating in over 180 countries worldwide.

- Responsible for providing technical input on statistical research methodologies that use quantitative data analysis.

- Developed statistical methods used in designing survey questionnaire for Child Activity Survey 2016, featuring data set of 25,000 households across all nine provinces of Sri Lanka and conducted by Department of Census and Statistics.

- Developed documentation on ILO statistical metadata collected through large-scale surveys in Sri Lanka.

Key achievements:

- Helped Programme Unit embrace a more research-based environment focused on ensuring the use of statistically robust methodology for studies requiring survey-based analysis.

Statistical Research Officer – Institute of Policy Studies (Colombo, Sri Lanka) **Jan 2011 – July 2014**

The IPS is the leading regional (South Asia) economic think-tank working closely with government, private sector, academia, and civil society to conduct policy-oriented economic research.

- Lead statistical researcher for study modeling food scarcity in Sri Lanka using multivariable regression and supply and price elasticities of key agricultural commodities for National Science Foundation.

- Lead statistical researcher for study modeling impact of migrant workers' remittances on real exchange rates in South Asian economies using multivariable regression.

Key achievements:

- Lifted maturity of junior research staff by advocating and ensuring statistically robust analytical approaches for quantitative research studies.

EDUCATION

MSc Data Science **Sep 2016 – Sep 2017**

King's College London, United Kingdom

- Dissertation: Used Python and Spark SQL to develop socio-demographic profiles of Transport for London tube passengers using geospatial analysis and k-means clustering on a dataset of 900 million passenger journeys.
- Coursework:
  - Designed schema and created relational database for a social network of employees using SQL.
  - Built forecast models to predict length of delays in international flights using multiple regression in R.
  - Created spam filter to classify over 5,600 SMS text messages using Scikit-learn and Natural Language Toolkit in Python.
  - Analyzed relationships between Twitter users and followers using MongoDB on dataset of 24,000 tweets.
  - Analyzed click-stream activity of website visitors using Hadoop MapReduce on dataset of 1 million records.

Master of Economics (Major in Econometrics) **Feb 2008 – June 2009**

University of Sydney, Australia

- Forecasted trends in GDP for Australia using monthly data from 1959 - 1995 comparing Simple Exponential Smoothing, Holt-Winters, and Box-Jenkins models.
- Analyzed Purchasing Power Parity trends incorporating error correction models, co-integration, and Dickey-Fuller testing.

Bachelor of Arts (Major in Economics) **Aug 2003 – June 2007**

University of Virginia, USA

PERSONAL PROJECTS **Dec 2019 – March 2020**

- Developing web scraper and text classifier for Twitter using Python that collects *n* most recent tweets for a specific hashtag during a specific period and classifies tweets based on sentiment and thematic area.
- Developing web scraper and text classifier for Reddit using Python that collects top comment (and other relevant details) from top *n* posts in a specific subreddit and classifies comments based on sentiment and thematic area.