

Kshitij Mamgain

Toronto, ON; Phone: 437-216-1818; Email: kshitij.mamgain@gmail.com

LinkedIn: <http://linkedin.com/in/kshitij-mamgain>; GitHub: <https://github.com/kshitijmamgain>

SUMMARY

- **4+ years of experience** in **data processing**, **database programming** and **data analytics**
- Efficient in utilizing tools of **exploratory data analysis** and **visualization** in **Tableau & PowerBI**
- Experienced in programming and statistical modeling using **Python**, **R** and **SAS**
- Hands on experience with **machine learning APIs** and **statistical learning models**
- Hands-on knowledge in **big data management** and **ETL development** using **Spark** and **Hive**
- Proficient in cloud computing using **GCP (Google Cloud Platform)** and **MS Azure**
- Excellent in **written and verbal interpersonal communication** and working with **agile** team
- Hands-on experience in **Object Oriented Programming (OOP)** and **Unit testing**
- Hands-on knowledge in container-based platform using **Docker**
- **Creative and innovative thinker, deep understanding of capital markets**
- Open for relocation
- Canadian Permanent Resident

TECHINICAL SKILLS

- **Programming Language** – Python, R, SAS
- **Data Tools** - MSSQL, MySQL, PostGres
- **Big Data Tools** - Hadoop, Hive, Spark, Dask
- **Version Control** - Git

RELEVANT EXPERIENCE

Research Assistant, Lantern Institute

Feb 2020 - Present

Supervisor: Tanaby Zibamanzar Mofrad

<https://github.com/kshitijmamgain/Higgs>

- Efficiently worked in **agile** team to **design, develop, and implement** end-to-end machine learning production pipelines on **Google cloud** on big data project with 11 million observations.
- Proficiently worked with **Dask and PySpark** libraries for big data exploration and **feature engineering**.
- Utilized solid understanding of Machine Learning, Statistical Analysis using Pandas, Numpy and Scikit-Learn to develop **scalable, repeatable, and secure data pipelines** for big data.
- Leveraged technologies like **Google AutoML** tables to compare big data and batch analytical solutions.
- Worked extensively with Pytest to code and test to maintain **code integrity** for robust applications.
- Designed methods to **collect, parse and visualize** large set of data with different design options.
- Expert in getting insights from complex algorithms like XGBoost and Catboost using SHAP
- Worked with peer developers to make sure that all data solutions are consistent and ensured all automated processes to preserve data

Data Science Fellow, Lantern Institute

Sep 2019 - Feb 2020

- **Samsung Health Data Analysis** - Project on behaviour analyses on health data. Performed data cleaning, munging to extract user information in Python using **numpy**, **pandas**, **datetime**, **matplotlib**, **seaborn** packages. <https://github.com/kshitijmamgain/Health-Analysis>
- **Statistical Analysis on Interest Rates** – Developed a project on forecasting and modeling LIBOR rates trends using advanced R packages like sde for solving differential equation, fitdistrplus for fitting distributions, ggplot for visualization. https://github.com/kshitijmamgain/R_Data_Analysis

Program Officer / Data Analyst , Tata Trusts

Jan 2015 - Jun 2019

- **Managed projects** on international development to define business requirements, KPIs and analytic solution design.
- Developed **analytical results** on social impacts through processing data sets from multiple sources.
- **Developed ETL** process and analyzed data through self-initiated projects using **SQL**
- **Collaborated with internal teams** and stakeholders to develop **statistical models**.
- **Built dashboards** and presentations using data visualization in **Tableau** for effective communication.
- Helped revamp the outreach reporting process and reduced **reporting turnaround by 80%**.

Solution Engineer, Nokia Solutions Network

Sep 2010 - Sep 2012

- Managed bids and solutions, bills of quantity and budgets, RFPs and RFI in telecom sales process.
- Provided technical solutions to the global customer teams on packet networks and data networking.
- Prepared networking solution and responses to technical business bids for telecom sector.

EDUCATION

- | | |
|---|----------------------------|
| • Certificate in Mathematical Finance & Data Science Program
Lantern Institute, Toronto | Sep 2019 - Feb 2020 |
| • Advanced Diploma in Data Science and Application
Metro College, Toronto | Jul 2019 - Mar 2020 |
| • Master's degree in Development Practice
Delhi, India (Evaluated by WES) | Jun 2012 - Jul 2014 |
| • Bachelor's degree in Engineering in Electronics and Communication
Delhi, India (Evaluated by WES) | Jul 2006 - Jun 2010 |

CERTIFICATIONS

- **Big Data, Smart Analytics, Machine Learning, and AI on GCP** from Coursera
Credential ID KP76JEBS2U7Y
- **Neural Networks and Deep Learning** from Coursera and Deeplearning.ai
Credential ID U4FGFGGKBGUQ
- **Improving Deep Neural Networks: Hyper-parameter tuning, Regularization and Optimization**
Credential ID 34GLYWYBVGZU
- **Sequence Model and Structuring Machine Learning Projects** from Coursera
Credential ID CXRUFP3PRC6S
- **Git Essential Training: The Basics** from Lynda.com
Credential ID EF6139578E304186ADB33A842AE988

PUBLICATIONS

<https://medium.com/@kshitij.mamgain>

- [Interest Rates Forecasting - using Stochastic Differential Equations](#) - a demo on modelling LIBOR interest rates using Vasicek model in R.
- [Leveraging combined power of SQL and Pandas](#) - a demo on utilizing pandas and sqlite package for data loading and wrangling in python.
- [Object-Oriented Programming to tune ML Model](#) - An illustration to create ML class for different hyper-parameter tuning techniques