Kshitij Mamgain

Toronto, ON; Phone: **437-216-1818**; Email: **kshitij.mamgain@gmail.com**
LinkedIn: **http://linkedin.com/in/kshitij-mamgain**

SUMMARY

5+ years of experience in data science and business analytics in finance and product development teams with strong knowledge of computer languages, database statistical and machine learning techniques

- 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**
- Open for relocation
- Canadian Permanent Resident

Technical Skills

- **Programming Language** – Python, R, SAS
- **Data Tools** - MSSQL, MySQL, PostGres, Pandas, Numpy, Sklearn
- **Big Data Tools** - Hadoop, Hive, Spark, SparkML, Dask
- **Analysis Tools** - Matplotlib Seaborn, Plotly, Pandas-Profiling, SHAP
- **Cloud Tools** - Azure, Google Cloud
- **Machine Learning** - TensorFlow, PyTorch, Keras, XGBoost, LightGBM, Catboost
- **Version Control** – Git, Google Source Repository

PROJECTS

Data Science Research Assistant,   **Feb 2020 - Present**
Lantern Institute, ON
UCI - Higgs Boson Data

- Efficiently worked in agile team to design, develop, and implement end-to-end data analysis on Google cloud platform on big data project with 11 million observations.
- Analyzed Higgs dataset in a group, developed Python packages for preprocessing, feature engineering, boosting methods (XGBoost, CatBoost, and Light GBM) and comparing different optimization methods such as Hyperopt, Optuna, and Random Search.
- 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 likeGoogle AutoMLtables to compare big data and batch analytical solutions.

Data Science Fellow,   **Jul 2019 - Feb 2020**
(Click on names to go to GitHub repo)
Samsung Health Data Analysis -- Python

- Project on behaviour analyses on real world data. Performed data cleaning, munging to extract user information and visualization in Python using **numpy**, **pandas**, **datetime**, **matplotlib**, **seaborn** packages.
- Created dynamic functions to analyze and visualize daily activity trends using function.
- Created modules to work with complex date and time features and standardize UTC zones.

[Statistical Analysis on Interest Rates](#) -- R

- 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.
- The codes included approaches to transform series into mean reverting and applying vasicek equation.

[Sentiment Analysis using NLP](#) -- Python

- Performed data analysis, visualization and A/B testing for customer reviews on different services.
- Developed and compared the results for different ML algorithms including SVM, Catboost.
- Achieved 85% accuracy on creating a model using Spark machine learning library.
- Used Pytorch for pre-trained BERT model to achieve 92% accuracy in sentiment prediction in Notebook.

[Credit Card Fraud Detection](#) -- Python

- Used highly imbalanced credit fraud dataset from Kaggle to create LightGBM machine learning model.
- Engineered minority class by utilizing and comparing SMOTE, Over-Sampling the minor class.
- Explored the model performance using auto encoders in tensorflow in an ANN model.
- Compared loss function and used Hyperopt to tune the model using Bayesian Optimization.

[Black Friday Sales](#) -- SAS

- Analyzed Black Friday sales data in SAS to identify the key business recommendation and customer classification by coding statistical procedures and backward feature elimination.

WORK EXPERIENCE

Program Analyst   **Jan 2017 - May 2019**
Tata Trusts, Delhi, India
- Managed projects to define business requirements, KPIs and analytic solution design.
- Developed analytical results on social impacts through data processing from multiple sources using SQL.
- Developed process and analyzed data by ensuring high data quality and integrity.
- Collaborated with internal teams and stakeholders to develop statistical forecasting models.
- Built dashboards and presentations using data visualization in Tableau for effective communication.
- Applied SAS procedures for advanced statistics like A/B testing, regression/GLM in the on-going projects.

Program Associate   **Oct 2014 - Dec 2016**
Tata Trusts, Mumbai, India
- Efficiently utilized multitasking skills to deliver under demanding timelines with multiple deliverables.
- Developed ETL process and analyzed data through self-initiated projects using SQL
- Competent in standard MS Office tools, such as Word, Outlook and Excel.
- Utilized analytical skills to work with budgets, MIS using SQL and undertake MS-Excel based modeling.
- Applied writing skills to prepare proposal writing, project pipelines and meeting briefs.
- Effectively communicated to interact with partners, government, and other stakeholders.

Solution Engineer   **Sep 2010 - Sep 2012**
Nokia Solutions Network, Delhi, India
- 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.
- Efficiently provided tendering support to market sales team as member of Global Sales Support.
- Prepared standard solutions and offered material (SoCs, Configuration, standard pricing).
- Matched customer needs with Nokia products/services and translated solutions to provide clear customer value and benefits.

EDUCATION

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

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 EF6139578E304186ADBAD33A842AE988

PUBLICATIONS   **https://medium.com/@kshitij.mamgain**

- **Object-Oriented Programming to tune ML Model** - An illustration to create ML class for different hyper-parameter tuning techniques
- **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.