RISHABH KUMAR KANDOI

Bachelor's in Computer Science & Engineering | Master's in Data Science

(585)-(410)-(8739) | rishabhk8@gmail.com | Digital Portfolio: http://rishabhkandoi.github.io

GitHub: https://github.com/RishabhKandoi | LinkedIn: https://www.linkedin.com/in/rishabh-kandoi/

CAREER OBJECTIVE

Team oriented individual with superb communication abilities seeks the position of a Data Scientist. Offering exceptional programming, statistical analysis, and research abilities to extract, clean, and prepare data for data science model ingestion.

WORK EXPERIENCE

- **PayTM Payments Bank,** leading Beneficiary Bank for UPI Transactions, India
  Senior Software Engineer *Sept 2021 – Aug 2022*
  - Improvised payment lifecycle for customers, with intuitive error handling, using Java with extensive focus on SQl query optimizations in order to work with more than 20 million transactions per day.
  - Resulted in 2% increase in user traffic on our android and IOS mobile apps.
- **BigBasket**, among the top 3 online retail in Food & Drink category, India
  Software Engineer *Aug 2019 – Aug 2021*
  - Developed 3 microservices and multiple functionalities in monolith, in Last Mile Delivery Flow.
  - Optimized algorithms and worked on cross-platforms to convert few heavy computations from Python to Spark (hosted on AWS) based approach, integrating flows between multiple microservices across multiple teams.
  - Estimated an increase in our cron jobs speed by 500%, with 100% more efficient results, by using Spark computations.
  - Incorporated Helm to automate deployment strategies, resulting in 50% faster production ready servers.

EDUCATION

- University of Rochester, NY, USA, Aug'22-May'23
  Master's in Data Science
- NIIT University, India – 3.9/4.0, 1st Rank Holder, 2015-19
  Bachelor's in CSE, with Data Science Specialization

TECHNICAL SKILLS

- Programming Languages: Beginner – C/C++, Java, RxJava, SAS || Proficient – Python, R.
- Database: Beginner – Hadoop, PostgreSQL || Intermediate – MongoDB, ElasticSearch || Proficient – MySQL.
- Infrastructure: Beginner – AWS, Jenkins || Intermediate – Kubernetes, Docker, Kafka
- Web Development: Beginner – HTML, JavaScript
- Tools: Beginner – Grafana, Redash, JIRA, Confluence, BitBucket, DataDog, Apptuit || Proficient – Kibana, GitHub, NewRelic
- Statistical Models: Beginner – A/B Testing, Generalized Linear Models, Tree-based Methods, Clustering & Mixture Models, Time Series Forecasting, Association Rules and Pattern Mining
- Data Science Packages: Beginner – SciPy, Scikit, TensorFlow, Statasmodels, PyTorch, Databricks
- Visualization: Beginner – Tableau, Plotly, ggplot || Advanced – Powerpoint, Excel

PROJECTS

- **Trauma Detection** – Applied exploratory data analysis, sampling, and ML modeling techniques to detect Trauma level for patients based upon some pre-defined lab tests and patient condition, to allot resources aptly.
- **Spam Mass Detection** – Detection of spam pages in search engines using Page Rank algorithm, exploiting MapReduce Framework in Hadoop (HDFS) for parallel processing of the big data.
- **Bike Rental Inventory Analytics** – Found meaningful demographic insights based upon historical data and performed time-series forecasting using real-time streaming data. Built ETL and MLFlow Pipelines, along with git integration on Databricks.
- **Effect of multiple socio-economic, and demographic factors along with mental health and major political events on crime rate within a city (Research Project)** – Correlation analysis of basic factors affecting crime rate, with our approach including sentiments from Twitter, and thus giving better prediction. Also identified the crime categories which have high count but population is not either aware of it or not vocal about it and thus needs awareness to eradicate such crimes.
- **Group Chat Text Segmentation Using Topic Modeling –** Segmented Slack Dataset using hierarchical Bayesian unsupervised topic segmentation model. Process involved Data cleaning, Tokenization, identifying reply objects, calculating similarity distance and naming the topic. Most prominent use-cases are decision auditing and dynamic responsibility allocation.

RELATED COURSES

- Time Series Analysis
- Intro to Statistical Machine Learning
- Data Mining
- Tools for Data Science
- Data at Scale
- Pricing Analytics