# DEEKSHA CHILUKURI

Toronto, ON | deekshachilukuri31@gmail.com | linkedin.deekshachilukuri | github./deekshachilukuri31

---

## EDUCATION

**MS in Data Science (Concentration Applied Machine Intelligence)**                     **Sep 2021- July 2023**

Northeastern University                                                                                   **Toronto, ON**

**Bachelor of Technology in Computer Science and Engineering**                         **Jun 2016 - Sep 2020**

Vignan's Institute of Information Technology                                                         **Vizag, India**

## SKILLS

- **Programming Languages:** R, **Python**, Scala, SAS, SQL, C, C++.
- **Frameworks & Libraries:** Pandas, NumPy, Pyspark, Scikit-learn, **NLTK**, **Pytorch**, Keras, **TensorFlow**, **mxnet, Hugging-face.**
- **Data Management:** MySQL, NoSQL, Big Data technologies (Apache Hadoop, Spark, Hue, Hive, **Databricks**)
- **Tools:** Alteryx, Looker, Tableau, R Studio**, Logstash**, GCP, Jira, GIT, Datadog, Docker, Azure / GCP , Kubernetes.
- **Machine Learning & Deep Learning**: Supervised & Unsupervised algorithms, Neural Networks (**CNN, RNN**), Generative Adversarial Networks (**GANs**), **Natural Language Processing (NLP)**, Reinforcement Learning and Computer Vision.

## EXPERIENCE

**Data Engineer**                                                                                        **Aug 2023 –  Present**

**Toronto-Dominion Bank (TD)**                                                                       **Toronto, ON**

- Orchestrated the migration of **12 diverse data sources** to **HDFS**, followed by seamless transfer to **Azure Data Lake.**
- Deployed code to advanced environments, demonstrating skillful monitoring of ETL loads through Azure Data Dog.
- Designed and implemented interactive Tableau dashboards to monitor ETL performance & maintain data quality standards.
- Generate large volumes of **Synthetic data** using Databricks Labs data generator for Robust Pipeline Validation.
- Configured **JIRA** for issue tracking and workflow collaboration as a part of an **Agile development** process.

**Data Analyst**                                                                                         **Sep 2022 – Jul 2023**

**Royal Bank of Canada (RBC)**                                                                       **Toronto, ON**

- Implemented a **SARIMA** time series model, **achieving 90% accuracy** in predicting workloads for resource allocation.
- Executed modeling of **holiday effects & seasonality** in weekly time series SARIMA model, resulting **11% surge** in accuracy.
- Optimized **API requests** for big data insertion into an **Elastic cluster**, reducing data latency from **1 day to 4 hours.**
- Used **SQL** to extracted data from databases and Implemented an ETL pipeline for **Near real time analytics**.
- Automated **Manual accessibility testing** on dashboards, resulting in a **50% reduction** in manual testing.
- Developed automated attrition analysis system using Python and SQL, reducing manual work by 80% and improving accuracy; generated **Tableau reports** for effective monthly attrition rate monitoring.

**Data Analyst**                                                                                         **Sep 2020 - Aug 2021**

**Essen Electronic Systems Pvt Ltd**                                                                 **Vizag, India**

- Conducted detailed analysis and generated insights and published reports using **Python, SQL, and Excel**.
- Develop and maintain **dashboards** and reports to track key metrics and communicate insights to stakeholders.
- Created visually appealing data visualizations and interactive dashboards using **Tableau** and **MS power BI.**
- Build an ETL pipeline using **PySpark** that ingested **1.5 million records** of data daily from 5 different data sources.
- Implemented a **web scrapper** in python that acquire data from different web applications.

**Data Science Intern**                                                                                  **May 2018 - Jul 2018**

**Verzeo**                                                                                               **Hyderabad, India**

- Prepared **10+** datasets for modeling in Python Via cleaning and preprocessing.
- Improved accuracy by **8%** using Stepwise variable selection & recursive feature elimination for **feature selection**.
- Formulated a model development plan by comparing the pros/cons of various Machine learning algorithms.

## PROJECTS

**Prediction of Emotional Tone of Songs using NLTK & PySpark**                    Mar 2023 - Mar 2023

- Developed a real-time emotion prediction system that allows users to input a URL of a Spotify song.
- Extracted the song & artist name from Spotify's API & retrieved the lyrics of the song from Musixmatch API.
- Used **NLP techniques** such as **VADER, BERT and GPT** to predict the emotion of the song based on its lyrics.

**Incremental Uplift Analysis for Promotional Campaigns**                    Feb 2023 - Feb 2023

- Analyzed the promotional campaigns' impact on sales by splitting the dataset into test groups and a control group.
- Identified 'discount offers' resulted in a higher uplift of **72 %,** compared to 'Buy One Get One' offer with a **43 %.**
- Validated results with statistical significance testing to ensure significant difference in uplift between the two offers.

**Sentiment Analysis on ChatGPT Tweets**                    Feb 2023 -Mar 2023

- Utilized **PySpark** to preprocess and large unstructured data containing **82129 tweets** with the hashtag #ChatGPT.
- Applied sentiment analysis techniques using **Bert** models to predict the sentiment of tweets.
- Visualized the sentiment analysis resulting in **15% positive** tweets, **78% neutral** tweets, and **7% negative** tweets.

**Toronto Real Estate Investment & Development**                    Apr 2022 - Apr 2022

- Built an Interactive Dashboard using **Tableau** to analyze Toronto Real Estate Board data.
- Provided insights about home price levels and trends in Toronto neighborhoods using **MLS® Home Price Index**.
- Identified **York Mills** as the most expensive region in Toronto, with an average price of around **1.8 million dollars**

## VOLUNTEER EXPERIENCE

**Volunteer at INDIA Redefined**                    Apr 2020-May 2020

- Volunteered for 4 weeks at an NGO and participated in Clean India and Green India initiatives.
- Organized awareness programs and provided free saplings, improving social skills through interactions with children.

**Student Coordinator at VISTA Tech Fest**                    Sep 2018 -Mar 2018

- Skillfully planned and executed the " Idea quest " student event, resulting in 600 students participating.
- Efficiently coordinated the "Wise Quackers" student event, ensuring seamless execution and a successful outcome.

## ACHIEVEMENTS

**Responsible AI 2023 Symposium**                    April 2023 -June 2023

- Awarded **first place** at the Responsible AI 2023 Symposium. Presented project on "Sensitive Information within Generative AI Chatbots". Demonstrated expertise in AI ethics, data security, and innovative problem-solving.

**Burst The Bug Challenge**                    Sep 2018 -Mar 2018

- Awarded **first place** in the Burst the Bug challenge hosted by Vignan's Institution of Information Technology, outperforming a field of 800 participants.