DIKSHANT JOSHI

- dikshant.joshi16@gmail.com
- www.linkedin.com/in/dikshant-joshi
- +1 (603)866-6044
- United States
- https://github.com/dikshant1603

EDUCATION

Masters of Science (Business Analytics) **University of New Hampshire, NH, USA** GPA: **3.83/4 |Graduation – 08/30/2023**

Courses: Business Intelligence, Probability and Simulation, Statistical learning, Communication of Data, Business Analytics using Python, Optimization Methods, Time Series Analysis, Big Data & AI

Bachelor of Engineering (Comp. Engg.) **G.B.P.U.A&T, Pantnagar, India** GPA: **7.084/10 | Graduation – 07/01/2021**

Courses: Data Structures & Algorithms, Database Management System, Artificial Intelligence, Data Mining & Warehousing

SKILLS

- **Languages:** SQL, Python, R, AMPL, JAVA, C, C++, HTML, CSS
- **Tools:** Power BI, R-Studio, R-Shiny, PostgreSQL, Tableau, Solver, RDS, Google-Colab, Jupyter, Alteryx, Data Robot, MySQL, Spark, Microsoft Word, Microsoft PowerPoint, Microsoft Excel, Project Management
- **Statistics:** Predictive Modelling, Unsupervised Learning, Machine Learning, NLP, Topic Modelling, Optimization, Web-Scraping, Hypothesis testing, Time Series Analysis, Big Data, Neural Network, TensorFlow, Causal forest, Statistical Inference

WORK EXPERIENCE

University of New Hampshire (Role: **Graduate Research Assistant**) **08/15/2022 – 07/30/2023**

- Analyzed large amounts of breach data for New Hampshire State, including cleansing, pre-processing, and text analysis. This involved breaking down the breach letter text into smaller units, such as words or phrases, and applying techniques such as stemming and stop-word removal using **Excel**, **Python** to improve the accuracy of the analysis
- Utilized BagofWords, and word2vec algorithms in **Python** to perform topic modeling on breach letters to identify common themes and patterns and know key issues and concerns raised by the breaches and understand the root causes of the problems
- Conducting sentiment analysis on the letters to assess the emotional tone of the breach letters, determine whether the letters were generally positive or negative in tone and identify any trends or patterns in the emotional content of the letters
- Superheaded an engaging Sports Analytics project as a Project Manager focused on Tennis ATP data, involving data cleansing, analysis, and visualization to extract valuable insights using MySQL workbench, Python, and Tableau. Delivered an engaging presentation to communicate the findings effectively.
- Utilized AWS to create RDS and connected data through MySQL Workbench for efficient database management

Accenture Solutions Pvt. Ltd. (Role: **Data Analyst)** **09/03/2021 – 04/21/2022**

- Collaborated with stakeholders for Business Planning identifying business opportunities and goals, to determine if analytical approach is suitable.
- For Healthcare Client developed solutions in **Python** and implemented an **ETL** process within a data warehouse and data integration project, featuring the establishment of an AWS RDS database. Leveraged **Apache Spark** for the implementation of a scalable big data solution, encompassing the acquisition, cleaning, and harmonization of data into a consistent scale.
- Applied data visualization techniques and built dashboards to analyze trend and seasonality using **R, Power BI, Tableau**, and built forecasting model to predict next quarter demands leading to **10%** improvement in the Accuracy using **Python, R** scripts
- Created Supply Chain Optimization model and leveraged underutilized data in product quality control leading to reduction in number of product recalls by **15%**
- Validated and tested models for adequacy and issues using Hypothesis Testing and Troubleshooting to support business objectives
- Communicated complex information clearly and concisely through Oral Communication using Presentation, partnered with cross functional teams to implement models, and lead changes in policies, procedures, and processes

PROJECTS

MNIST-Image classification **03/2023 – 05/2023**

- Applied data preprocessing techniques, including reshaping images and converting labels to categorical format, to prepare the MNIST dataset for training.
- Developed and implemented a CNN for image classification, achieving **98%** accuracy in recognizing handwritten digits.

Airbnb **10/2022 – 12/2022**

- Cleansed/organized the data using **Power Query** to create visualizations, correlation matrix and interactive dashboard in **Power BI** showcasing listings, review ratings, and demographic prices for Boston
- Conducted a causal inference analysis, unveiling an **8.7%** increase in booking rates post-Airbnb's Experiences & Trips Program launched in November 2016. Utilized PC algorithm to create CPDAG for a comprehensive understanding of data relationships.
- Employed causal forest models to analyze Airbnb's complaint line and safety measures, revealing location-specific insights, including a nuanced **6%** booking rate increase post-safety measures in November 2019.

E-Commerce Data – Delivery Prediction **01/2023 – 03/2023**

- Cleansed/Organized data in **R** and performed **EDA**.
- Utilized Logistic regression, K-Nearest Neighbors (KNN), XG-Boost, and tree-based models to predict On-time delivery (Yes/No) attaining a high Recall of **92%**, AUC of **0.72** enhancing customer satisfaction.

Stocks – Time Series Forecasting **03/2023 – 05/2023**

- Curated and structured stock data for Indian Car Manufacturers to make the series stationary.
- Applied time series models (MLR, ARIMA, ARCH/GARCH) to forecast one-year closing prices for Tata. Evaluated model using AIC and MAPE, selecting GARCH with minimum **AIC (6698)** and **MAPE (5%)** for accurate predictions on historical data.

Spotify – Cluster Analysis **03/2023 – 05/2023**

- Cleansed and organized Spotify data using R, producing meaningful visualizations for EDA. Employed advanced techniques such as K-means and Mixture Model to cluster Spotify songs based on their characteristics, enhancing the depth of the analytical insights.

Optimization **01/2023 – 03/2023**

- Created an Optimization model using **AMPL IDE** and **CPLEX** Solver to optimize and give shortest route solving Travelling Salesman Problem
- Created an Optimization model using **CVXPY** library in Python and **GLPKMI** Solver to optimize and give minimum number of colors required to color nodes of a graph solving Graph Coloring Problem

COVID-19 Vaccination Awareness and Aftermath **10/2022 – 12/2022**

- Cleansed and prepared the Covid-19 data of the United States using **Power Query** tool of **Excel, R** to perform **EDA**
- Developed **Tableau** dashboards illustrating Covid-19 trends from 2019-Present and US Vaccine Hesitancy since early 2021.
- Scraped Covid-19 twitter data and conducted sentiment analysis on tweets using R, visualized results in **Power BI**.

Data Wrangling and Analysis using Yelp data **10/2022 – 12/2022**

- Pre-processed data in **R** using tidyverse and tidytext to make it analysis ready.
- Utilized R's ggplot, ggplot2, and leaflet libraries for creating visualizations, maps, and conducting Sentiment Analysis on hotel reviews. Additionally, developed and published an interactive dashboard using **R Shiny**.

CERTIFICATIONS

- NIIT Certification of OOP using C++.
- Coursera Certification of IBM Data Science Professional Certificate.