DIKSHANT JOSHI

• dikshant.joshi16@gmail.com • www.linkedin.com/in/dikshant-joshi • +1 (603)866-6044 • United States
• (Dikshant Joshi) · GitHub

EDUCATION

Masters of Science (Business Analytics) **University of New Hampshire, NH, USA** GPA: **3.83/4 |Graduation – 08/2023**
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 – 07/2021**
Courses: Data Structures & Algorithms, Database Management System, Artificial Intelligence, Data Mining & Warehousing

SKILLS

- **Languages:** SQL, Python, R, AMPL, JAVA, C, C++, HTML, CSS, SAS, SAP
- **Tools:** Power BI, R-Studio, Excel (Pivot table, Lookups, Solver), PostgreSQL, Tableau, RDS, Jupyter, Alteryx, Data Robot, MySQL
- **Statistics:** Regression & Classification Techniques, Unsupervised Learning, Natural Language Processing (NLP), Topic Modelling, Optimization, Amazon Web Services, Web-Scraping, Hypothesis testing, Trend Analysis, Time Series Analysis, Big Data, Neural Network, TensorFlow, Causal Forest, Deep learning, Machine learning, Image Analysis

WORK EXPERIENCE

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

- Spearheaded the preprocessing and data modeling of extensive breach data for the New Hampshire State, employing advanced techniques such as tokenization, stemming, and stop-word removal using **Python** to elevate the accuracy of the analysis.
- Leveraged BagofWords and word2vec algorithms in **Python** to conduct sophisticated topic modeling on preprocessed breach letters identifying common themes and patterns, facilitating root cause analysis of key issues raised by breaches.
- Conducted 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
- Utilized **AWS** to create Relational Databases and connected data through **MySQL** Workbench for efficient data management
- Led an engaging Sports Analytics project focused on Tennis ATP data, involving extraction, cleansing and data analysis to gain valuable insights using **MySQL workbench**, **Excel**, and **Tableau**. Conducted in-depth case study of BNP Paribas 2013 Paris Masters match between Rafael Nadal and David Ferrer(winner), revealing a misjudgment in higher betting odds for David Ferrer (6.5).

Accenture Solutions Pvt. Ltd. (Role: **Data Analyst)** **07/2021 – 07/2022**

- Collaborated with stakeholders to document business processes and gather requirements. Developed and implemented Python scripts to create data pipelines, streamlining data processing workflows resulting in **25%** reduction in deployment time.
- For Healthcare Client Engineered and optimized Extract, Transform, Load (**ETL**) processes, utilizing Informatica as the ETL tool and leveraging GCP as the cloud data warehouse for scalable data verification, cleansing, and harmonization ensuring data quality.
- Utilized Advanced **Excel** and **SQL** for data modeling, focusing on Online Analytical Processing (OLAP), to design and implement dimensional models like star and snowflake schemas. Crafted **SQL** queries to create and optimize OLAP cubes, aggregations, and hierarchies, enabling efficient analysis and reporting for decision-making
- Utilized Time Series to analyze trend and seasonality crafting a forecasting model that accurately predicted quarterly demands for Medical supplies using **Python** and **R** scripts, resulting in a notable **10%** enhancement in forecasting accuracy.
- Utilized root cause analysis techniques to investigate quality control issues and implemented an optimization model based on root cause findings, resulting in a **15%** reduction in the number of product recalls. Conducted Hypothesis testing to assess models.
- Collaborated with stakeholders to define comprehensive test plans and conducted User Acceptance Testing (**UAT**) to ensure that analytical models aligned seamlessly with business objectives, thus facilitating smooth implementation and adoption.
- Utilized visualization tool like **Tableau, Power BI** to build dashboards and reports, communicating complex information and KPIS with clear objectives effectively and concisely to diverse stakeholders. Worked with cross functional trams and actively participated in backlog review and Sprint planning sessions, following **Agile** methodology

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**

- Ensured Quality Assurance by Cleansing/data validation of 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%** increase in booking rate 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. Utilized Alteryx predictive modeling to further validate the results by creating workflows.

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.

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 **Microsoft 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 and Python.