

# Nikunj Malpani

[www.nikunjmalpani.com](http://www.nikunjmalpani.com) | Email ID : [malpaninikunj78@gmail.com](mailto:malpaninikunj78@gmail.com) | Mobile No.: +1-424-420-8337

Linkedin: <https://www.linkedin.com/in/nikunj-malpani-59555811a/>

## EDUCATION

**INDIANA UNIVERSITY**, Luddy School of Informatics **Bloomington, INDIANA**  
**Master of Science in Data Science, CGPA: 3.83 / 4.0**  
**Graduate Associate Instructor for Statistics - S520** Aug 2021 - May 2023  
**SVVV UNIVERSITY**, Department of Computer Science Engineering **Indore, INDIA**  
**Bachelor of Technology in Computer Science & Engineering** Jul 2016 - May 2020

**CERTIFICATION: AWS Certified Cloud Practitioner (CLF-C02)**

## WORK EXPERIENCES

**St. Jude Children's Research Hospital** | *Associate Biostatistician* June 2023 - Present

- Conducted **survival analysis** on clinical trials using **R** to investigate the effect of predictor on hazard rate.
- Innovated an auto data querying program for data validation, ensuring reliability of research data before storing in MySQL, while saving **300 hours/year**.
- Providing statistical support to **5+ clinical studies** focusing on pediatric cancer by utilizing statistical methods, interpreting statistical results and making valid inferences.

**Amazon** | *Data Scientist Intern* May 2022 - Aug 2022

- Built end-to-end ML project by developing an **attribution model** on **1B+** customer records using AWS services like Redshift, Sagemaker, EC2, S3 to correct bias in regional and glance view in-stock based metrics, leading to **improved product placement and enhanced customer experience**.
- Quantified the difference between two metrics caused by combination of data dimensions thereby informing the areas of improvement Amazon should focus on in the whole supply chain journey.

**Indiana University** | *Data Analyst* Nov 2021 - May 2022

- Engineered data pipelines to stream data from university's data warehouses, **reducing data delivery time by 40%**, thus automating the dashboard creation in Tableau to find the trend of company's alliance with the university.
- Examined the top contributing companies to our organization and made changes in the scoring criterion for the best operational approach, leading to the **implementation of a project development plan for 5 campuses**.

**Paytm** | *Business Analyst* Jul 2020 - Apr 2021

- Conceptualized reports and dashboards in Tableau for **5+ teams** for monitoring daily operations and performance.
- Devised a macro to automate the daily task of sending reports to merchants thereby reducing the time taken from **3 hours daily to 5 minutes**, achieving full accuracy while saving **750 hours/year**, using Visual Basic Analysis.
- Established the step-by-step process for reconciliation of inventory within Fulfilment Centres to offset monthly losses against Paytm to the tune of **INR 50 Million/year (600k USD)**.
- Evaluated the difference between the courier delivery Turn Around Times(TATs) and the Paytm benchmarks using complex SQL queries and rectified the same in the system which **decreased the customer cancellations by 20%**.

**LT Foods Ltd.** | *Junior Data Scientist* Jan 2020 - May 2020

- Analyzed marketing campaigns using Test Control methodology and incremental study; by generating key insights and refining future campaigns, which **increased revenue by 5%**.
- Performed predictive analysis using statistical methods and synthesized trend data to develop futuristic strategies.

**CJBS, University Of Cambridge** | *Data Science Intern, Prof. Jaideep Prabhu* Jun 2019 - Aug 2019

- Conducted ontological predictions for sentimental analysis on the data of **behavioral study of 200 college students**.
- Used R to run Analysis of Variance (ANOVA) test and Linear Regression with LIWC Results and found 3 most important variables which affected the study.

## SKILLS

- **Programming** : SQL, Python(Pandas, Numpy, Tensorflow, Scikit, Pytorch), R, VBA, Javascript, PHP, HTML, CSS
- **Tools and Miscellaneous Technologies**: AWS Data Pipeline (EC2, Lambda, Athena, Glue(ETL), S3, Sagemaker, Redshift, Datanet), Tableau, Power BI, Hive, Apache Spark, Docker, Git, Flask, MongoDB, MySQL, Snowflake, Oracle
- **Data Science, Big Data and Machine Learning**: A/B Testing, ETL, Principal Component Analysis, Regression, Naive Bayes, SVM, Gradient Boosted Models, LSTM, NLP, Time Series Forecast, ANN, CNN, RNN, Self Organizing Maps, Autoencoders, Boltzmann Machine, Hadoop, ANOVA, Generative Adversarial Networks

## PROJECTS

**Phishing Email Classification** | *Python / NLP*

- Wrangled meta data and analyzed email content to engineer multidimensional features via Named Entity Recognition.
- Developed XGBoost model with hyper-parameter tuning using 10-fold CV achieving the **AUC-ROC score of 0.90**.

**Dog Generative Adversarial Network (GAN)** | *Pytorch / Neural Network*

- Developed a PyTorch-based dog GAN project, utilizing state-of-the-art deep learning techniques to create realistic dog images through Generative Adversarial Networks.

**Twitter Football Data Analysis** | *Big Data / Cloud Computing / AWS Data Pipeline*

- Designed and implemented a real-time data pipeline to process unstructured data from Twitter on AWS platform.
- Used EC2 for scheduling, Glue for ETL, Sagemaker to model and Quicksight to visualize the real-time analysis.