

ROHIT GHOSH

+1 (860) 897-6881 | rohitghoshwork24@gmail.com | [LinkedIn](#)

SUMMARY

Experienced Data Scientist with a Master's in Business Analytics and Project Management, bringing 3+ years of expertise in developing data-intensive applications. Specialized in customer segmentation, skilled in predictive modeling, data processing, and data mining using Python, R, SQL, and adept at visualization with Power BI. Translates business needs into impactful deliverables.

PROFESSIONAL EXPERIENCE

Bensussen Deutsch & Associates, LLC. | Data Science Intern, Washington, USA **June 2023 – August 2023**

- Implemented Gaussian Mixture Model using R and Python for customer segmentation to categorize 200,000 customers into 4 dynamic segments with an 89% accuracy, leading to improvements in strategies for customer engagement.
- Executed intricate SQL queries within Snowflake, unveiling crucial insights into customer journey analytics and touchpoints, contributing to a 25% improvement in the accuracy of purchase pattern analysis, optimizing marketing efforts and resource allocation.
- Leveraged CLVTools package in R to build a Customer Lifetime Value model, integrating 6 significant covariates, achieving an 89% accuracy rate, providing actionable insights into future customer behaviour and transaction dynamics.
- Pioneered the development and application of fuzzy matching algorithm model for detection and elimination of duplicate entries within item and vendor datasets with a 96% overall match rate, resulting in improved data quality, and minimizing redundancy.
- Implemented advanced anomaly detection using Scikit-learn and Pandas, enhancing data integrity and fortified the reliability of analytical models, contributing to more informed decision-making processes for marketing and operations stakeholders.

Noteshelf (Fluid Touch Pvt. Ltd.) | Business Analyst – User Experience, Hyderabad, India **August 2020 – February 2022**

- Administered dashboards using Excel, Power BI, and Tableau to analyse, visualize, and document important user insights and KPIs about the Noteshelf app, resulting in a streamlined flow of information across teams and a 20% increase in data accessibility.
- Developed and implemented a new data validation system that improved data accuracy by 35% and reduced errors by 28%, resulting in more reliable insights for decision-making for marketing and android development teams of Noteshelf – Android.
- Collaborated with a team of software developers to identify and implement 200+ unique feature requests, 50+ UI/UX suggestions, and 600+ support issues across 15 production versions of Noteshelf – Android, improving user ratings by 35% in Google Play Store.

Think and Learn Pvt. Ltd. | Data Analyst, Kolkata, India **June 2019 – May 2020**

- Orchestrated the implementation time-series forecasting using Python to enhance sales forecasting accuracy by 15%, resulting in improvement in account selection accuracy by 40% and facilitating revenue generation of \$250,000 annually.
- Leveraged Power BI to build dashboards, yielding tangible results by analysing previous quarters sales data, yielding a 20% increase in sales team efficiency enabling real-time monitoring of KPIs, fostering a 15% improvement in decision-making accuracy.
- Led the sales team in data-driven initiatives using SQL and optimized pipelines, resulting in a substantial 25% revenue boost from targeted efforts. Extracted actionable insights to drive strategic decision-making processes.

ACADEMIC PROJECTS

Time Series Forecasting for National Electricity Demand of UK | SAS Studio | Excel | Python

- Conducted time series exploration on national electricity demand data using Python, Excel and SAS Studio, incorporating wind and solar power generation as independent variables, to develop forecasting models.
- Analysed ACF, PACF, IACF, and White Noise graphs to determine optimal parameters for Exponential Smoothing and ARIMA models, achieving a 98% accurate forecast for national demand, following rigorous statistical model comparison and selection.

Portfolio Allocation Analysis | Python (PySpark, Pyomo, yahoofinance)

- Crafted a diverse portfolio comprising 9 stocks, seamlessly integrating risk datasets using NumPy and Pandas techniques in Python.
- Executed training on 5 years of historical stock data and utilized MPT and moving averages algorithm to find best trading strategy out of Buy and Hold, Buy and Sell and S&P 500 with buy and sell yielding a profit of 28%.

Insurance Fraud Detection | Python (NumPy, Pandas, Matplotlib)

- Built machine learning models on Python with hyperparameters, using cross-validation to select the optimal fraud detection model.
- Achieved a F1 score of 0.82 by adjusting the classification threshold to optimize results on test and training data.
- Utilized permutation importance and created partial dependence plots to identify the top five influential variables in causing policy cancellations, providing actionable business recommendations based on the model's insights.

Bounty Burgers Decision Support System | R | R Shiny

- Spearheaded decision support system development using R Shiny package for Burger Bounty, creating a user-friendly Shiny app for data entry and implementing a dynamic dashboard to visualize aggregate sales, optimizing food truck operations.
- Developed a user-friendly Shiny app with the UI having 17 interactive input variables and the Server containing algorithm to read historical data and run 6 linear multiple regression models to estimate burger revenue and facilitate location recommendations.

EDUCATION

University of Connecticut School of Business - Hartford, CT

August 2022 - May 2024

Master of Business Analytics and Project Management (Business Data Science Concentration)

CGPA: 3.71/4.0

SKILLS

Programming languages: SQL, Python, R, SQL, C, C++, shell scripting

Database: Snowflake, MariaDB, Teradata, MySQL, SQL Server, DBeaver, PostgreSQL, Oracle, NoSQL, Redshift, DynamoDB

Cloud, ETL and Analytics: AWS, GCP, Tableau, Power BI, Advanced Excel, JMP (SAS), SAS Enterprise Miner, SAS Studio

Big Data: PySpark, Databricks, Hadoop (HIVE, HDFS, Map Reduce)

Design tools and Methodologies: JIRA, Confluence, Visio, Agile, Scrum, Waterfall, Gantt Chart, Waterfall, NLP

Coursework: Data Mining, Data Science, Business Statistics using R, Predictive Modeling, Adaptive Business Intelligence, Visual Analytics