

## Akanksha Sharma

+1 206-591-9425 | [akankshasharmadid@gmail.com](mailto:akankshasharmadid@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Leetcode](#)

Data Scientist with 4 years of experience in building and deploying ML models and data engineering/ETL pipelines with various tools and technologies, including Python, SQL, and cloud platforms. Proficient in implementing data-driven solutions to real-world business problems using Statistical analysis, Data visualization, and Data management

### KEY SKILLS

**Data Science:** Proficient in building Machine Learning pipelines, Chatbots and experienced in working on Time-Series data, Clustering Algorithms, Recommender system using Deep Neural Network and Natural Language Processing

**Data Engineering:** Design and implementation of scalable data pipelines for large volume of data (>10MM records), using AWS Kinesis & Lambda; Data migration automation including automation of data cleaning, validation, deduplication, wrangling & profiling

**Data Visualization:** Designing dashboards to track KPIs with user interaction, and business-specific filters & parameters

**User applications:** Development of web-based UI using CSS, PHP and chatbots to automate routine customer service tasks

**Testing:** Experienced in A/B testing and data analysis to optimize website performance and user experience

**Programming and Databases:** Python, Tensorflow, Keras, Scikit-Learn, R, SQL, MongoDB, PostgreSQL, Hadoop, PySpark

**Machine Learning:** Natural Language Understanding (NLU), Regression Analysis, Sentiment Analysis, Random Forests, BERT, CNN, RNN, LSTM, Classification, Clustering, AdaBoost, Decision Trees, RASA, Folium, Feature Engineering

**Domains And Focus Area:** Natural Language Processing (NLP), Operational Analytics, Recommender Systems, Predictive Analytics, Database Systems, Data Mining, Data Analytics, AI/ML, Deep Learning, Statistical Analysis, Permutation Testing, Hypothesis Testing,

**Tools and Technologies:** Git, Amazon Web Services, LaTeX, PyCharm, Google Analytics, Tableau, PowerBI, MATLAB

### EXPERIENCE

#### Syniti – Senior Data Consultant

January 2020 – August 2022

- Developed an automated SQL Query Generator using Python which consumes data requirements via an UI, eliminating the dependency on BI team for data analyses/reporting requirements, saving at least 1.25 BI head count per year
- Developed a customer support chatbot by eliminating Customer wait time and reducing resources by 80% (from 10 to 2 HC/year)
- Implemented a real-time support ticket routing system using Logistic Regression and Decision Trees, reducing response time by 95% (from 20 to 1 min/ticket) with 98% accuracy of ticket classification based on severity
- Owned and led new hire technical training and demo project mentoring. Trained 40+ new hires in a span of six months

#### Wipro – Data Engineer

July 2018 – January 2020

- Led the data migration from Mainframe to HANA by creating structured de-normalized ERP (Enterprise Resource Planning) datasets to be used for KPI measurement and automated reporting, reducing manual effort by 93% (from 12 to 1 HC/year)
- Automate data cleaning, validation and data quality checks during data migration using Regex, Dedup and Pandas, decreasing the processing time by 97% (from 100 to 4 hrs) for each data migration cycle, consisting of at least 1B+ data records
- Developed executive-level project status Tableau dashboards for milestone status, blockers/risks, and call to action focus areas

#### Grozip - Intern

December 2016 – January 2017

- Designed and launched a chatbot on the Facebook Messenger platform for real-time customer support and product search functionality, improving referral traffic by 54% (from 130 users to 200 users per day)
- Led the redesign of data lake architecture for new web application during migration, using an Entity-Relationship (ER) diagram resulting in efficient data retrieval and reduced deduplication by 75% (from 4MM to 1MM records)
- Designed a targeted campaign framework, using Google Analytics, to identify focus regions based on website traffic and user behavior for sales and marketing teams, increasing revenue-per-sale by 60% (9 to 14B per year)

### PROJECTS

#### Seattle Crash Data Severity Predictor <https://github.com/seattlelizzard/CrashSeattle>

- Developed a Crash Severity Predictor using Random Forest and Logistic Regression (Lasso) models using location, vehicle density, light, weather, and road conditions as features. Improved the accuracy of the model to 98% from 76% by (1) imputing 10% missing data using FancyImputer, KNN Imputer and (2) K-cross validation and hyperparameter tuning

#### NYC Flights Data Set Analysis <https://github.com/akankshasharmadid/R-nycflights13-UA>

- Performed statistical Time Series Analysis, Hypothesis testing, Bootstrapped permutation test on features like time, weather, precipitation using Tidyverse and ggplot libraries to analyze the departure delay and relative gain of a flights

### EDUCATION

University	Course	CGPA	Relevant Coursework
Seattle University	MS in Data Science (September 2022 – Present)	3.95 / 4 (2x Dean's Honor List)	Statistical Inference & ML, Data Science, Data Visualization, Text Processing & Searching, Big Data
VIT University	B.Tech - Computer Science and Engineering (2014-2018)	8.90 / 10	Cloud Computing, Software Engineering, Data Structures & Algorithms, Algorithm Design, Agent based intelligent systems