

# MONIKA BAGYAL

Fort Mill, SC | P: 201-210-9083 | [monika.bagyal@gmail.com](mailto:monika.bagyal@gmail.com) | [linkedin.com/in/mbagyal](https://www.linkedin.com/in/mbagyal) |  
Website - <https://minsifye.github.io/portfolio/>

Software Engineer with 6-years of work experience building data solutions and solving complex data problems on large datasets. Aiming to leverage my diverse skillset to solve challenging problems and adding business growth.

## EDUCATION

---

### University of North Carolina at Charlotte

Charlotte, NC, Dec 2020

Master's in Computer Science – Data Science Concentration

### Birla Institute of Technology & Science

Rajasthan, India, Nov 2014

Master's in Software Engineering

### HNB Garhwal University

Uttarakhand, India, Jun 2010

Bachelor of Computer Applications

## TECHNICAL SKILLS

---

**Programming:** Python, SQL, PL/SQL, ETL pipelines, PySpark, Unix Shell Scripting, Pandas, Numpy, Scikit-learn, pytorch.

**Tools:** Tableau, Oracle, AWS Spark, S3, EMR, Git, MS SQL Server, Data Warehousing, Informatica, Autosys, Actimize.

**Coursework:** Machine Learning, NLP, Computer Vision, Algo & DS, Cloud Computing, Visual Analytics, Big Data Analytics.

## WORK EXPERIENCE

---

### Wipro – Client UBS Bank, Bangalore, India

Senior Software Engineer, Nov 2016 – May 2017

- Designed and developed a data analysis project to provide a statistical report on payment data using changeable threshold.
- Automated a process of scheduling ETL pipelines to load 12 months of customer profiling records in data warehousing prod env.
- Supervised a team of 2 software engineers while ensuring to follow best practices of SDLC across multiple projects.

### BNP Paribas, Mumbai, India

Senior Software Engineer, Dec 2014 – Oct 2015

- Built ETL pipelines to process 10+ millions of trades records, using data filtering, data manipulation and data transformation.
- Created a complex Scoring Algorithm to assign a risk score between 0 to 100, based on highly risked client's trade activity.
- Optimized existing models in production by refactoring code to ingest multiple regions data, documented the changes.
- Experienced in handling production issues, skilled in communicating with technical and non-technical groups.

### Wipro – Client UBS Bank, Bangalore, India

Software Engineer, Jul 2010 – Dec 2014

- Reduced project cost, by developing a solution to read unstructured data file (approx. 5 TB of data) using Shell Scripting.
- Optimized Informatica ETL mappings to improve performance and reduce data redundancy using data modeling concepts.
- Audited authorized user activity and thresholds changes in database using PL/SQL to report and keep system compliant.
- Implemented a Risk score filtering approach using SQL based on customer's KYC and historical data to produce new alerts.

## ACADEMIC PROJECTS

---

### Sparkify - Data Science Project on AWS Spark, Apr 2020

- Performed data analysis and feature engineering using pySpark on music streaming data for customer churn problem.
- Trained a model using Gradient boosting classifier to achieve a higher F1 score on AWS Spark to process 12GB of data.

### Disaster Response Model, Mar 2020

- Analyzed disaster data to build a model for an API that classifies 26000 disaster messages. Built a webapp using Flask.
- Build an ETL & ML pipeline, to perform text processing using NLP and load it into SQLite, to predict 36 disaster categories.

### Recommendations with IBM, Mar 2020

- Generated individual article recommendations for similar users by using collaborative filtering approach.
- Predicted user ratings by applying SVD to evaluate overall performance for new recommendations.

### Hackathon - NLP project on Yelp Reviews Data, Feb 2020

- Project idea was to provide key insights from reviews to help customer make decision faster.
- Generated visualizations after text cleaning, normalization, tokenization, generated bag-of-words etc.
- Presented this project to 300+ people during All Women Hackathon (Pearl Hacks) at UNC Chapel-hill.

### Identify Customer Segments, Apr 2019

- Applied PCA to reach 99% of variance with 74% of features, after data cleaning, transformation and feature engineering.
- Discovered 14 clusters (K-means), to find target audience to achieve highest rate of returns from direct marketing.

## Certifications

- Data Scientist Nanodegree by Udacity (Apr 2020).
- Tableau 10 A-Z: Hands-On Tableau Training for Data Science on Udemy (May 2018).
- AML and KYC by Indian Institute of Banking & Finance (Oct 2013).