

BHUSHAN PATIL

Bhushan.pat38@gmail.com | Bloomington, IN | [LinkedIn](#) | [Github](#) | [Kaggle](#)

EDUCATION

Indiana University - Bloomington, IN
Master of Science (MS) in Data Science

Jan 2021- May 2022
CGPA: 3.1/4.00

WORK EXPERIENCE

Serve IT

Data Science Consultant

Bloomington, IN
Aug 2021-Ongoing

- Working on risk and fraud detection data and present insights and communicate it to the stakeholders.
- Created data warehouses in AWS Redshift and AirTable for inventory management to reduce costs by **\$10,000 annually**.
- Define and develop the program for feature engineering, data cleaning, manipulation to feed it to the model for training.

ZS Associates

Data Analyst

Pune, India
Mar 2019 - Jul 2021

- Led the analytics and business operations team to review and design strategy for client meetings and communications.
- Performed A/B testing, Z-tests to make decisions for launching new features on application, increasing app traffic by **36%**.
- Deployed a recommendation engine using Natural Language Processing in AWS Sagemaker for reps to pitch drugs based on doctor's past order history to **increase** Quarter-end sales by **16%**
- Using predictive modeling, created rep alignment for new drug launch to **increase revenue** to **\$56k** from expected **\$40k**
- Deployed forecasting methods on historic data and used to collaborate with client by setting goals for the reps
- Optimized machine learning operations to reduce redundancy, manual effort, and process time by **50%**.
- Created Data Warehouses and ETL systems for several pharma clients in AWS Redshift, AWS Glue and Apache Airflow.
- Migrated project framework from Informatica and Oracle to AWS Redshift and EC2, saving **18%** in **operations costs**
- Built Weekly Sales reporting in Tableau and MicroStrategy for a team of ~1000 reps.

Atos

Software Developer

Bangalore, India
Sep 2018-Mar 2019

- Created **~15** innovative automation tools in Python to save **~3 hours** of team's effort daily.
- Used Exploratory data analysis and statistical modeling for problem solving and to convey insights to the client.
- Owned the development of data systems and processes for effective utilization, identified continual areas of improvement.

PROJECTS

Healthy Harvest – Python, Convolutional Neural Network, Keras Callback, Tensorflow, Machine Learning, Deep Learning

- Using CNN layers and Keras callback functions, the program identifies diseased and infected plants from image data. Validated early stopping, model checkpoint and reduced learning rate functions for accuracy of **98.90%** over 40 epochs.

Digits Recognizer – Python, Convolutional Neural Network, Tensorflow, Machine Learning, Deep Learning

- Implemented a system identify handwritten images of numbers using CNN layers with 'reLu' as the activation function.
- Using data augmentation techniques, which zooms, rotates, and whitens the image, the model was able to differentiate between 7 and 9 which were causing issues earlier. Data augmentation boosts the accuracy from **98.5%** to **99.32%**

Automated Paper Grading System – Python, Natural Language Processing, Wordnet, NLTK

- Developed using Natural Language Processing and Semantic Analysis in Python to give us an accuracy of **96.34%**. Using semantic analysis and Ontology, the similarity index and score of student's and ideal answer is calculated.
- Published research paper 'Automated Paper Grading System' in [IEEE explorer](#) and presented in [ICICCT conference](#)

Fantasy Football – Python, Exploratory Data Analysis, Support Vector Regression, Decision Trees, Dask, SciKit Learn

- Performed EDA to identify patterns, key metrics, challenges, and trends of players performance
- Achieved an accuracy of **94.26%** by training regression models in Python and used cross-validation experiments to predict the players performance in a season based on last year's data.

TECHNICAL SKILLS

Programming – Python, Java, R, SAS, Numpy, Pandas, Scipy, Sci-kit Learn, Tensorflow, Matplotlib

Database Tools – SQL, PostgreSQL, AWS Redshift, MongoDB, Hive

Skills – Deep Learning, Statistics, Hypothesis Testing, User segmentation, A/B Testing, Sampling, Docker, Github

Amazon web Services – S3, EC2, Redshift, Kinesis, Glue, Sagemaker, Athena, Aurora, DynamoDB, Lambda

Data Visualization – Tableau, MicroStrategy, Power BI, Looker

Big Data – Pig, Kafka, Spark, Hadoop, Map Reduce, PySpark

Machine Learning – Random Forest, Regression, Decision trees, Recommendation systems, Time-series analysis, Conjoint analysis, Gradient Boosting, AdaBoost

Tools/IDE's – Jupyter, Anaconda, RStudio, Databricks

ACHIEVEMENTS

- Secured 3rd place at GT-IDEA Case competition organized by Grant Thornton in Indiana University 2021
- Professional Certified AWS Cloud Practitioner 2021