Anuj Kulkarni

Data Scientist

✉ avkul2601@gmail.com   📞 8530011914   in https://www.linkedin.com/in/anuj-kulkarni-954b63221

Profile

An Analytics professional Holding 4.5 years of Experience with Proficiency in Machine learning, Big Data Analytics, statistical modeling in Descriptive Statistics, Forecasting/Predictive Analytics, Segmentation Methodologies, Regression Based Models, Hypothesis Testing, Time series, Random Forest. Rich experience of building user friendly reports to give better visibility of business and aid Strategy & Planning. Experience in statistical tools and a keen communicator with strong problem solving, planning, presentation and people management skills.

Skills

Big Data (Hive,Impala,Hbase,Kafka,Spark)   **Data Science** (EDA,Feature engineering,Feature selection)

Machine Learning (Classification,Regression,Clustering,Time-series)

Python (Pandas,Numpy,Matplotlib,Scikit-learn,Seaborn,Tensorflow)

AWS (EC2,Sagemaker,Glue,S3,lambada,Cloudwatch)   **SQL**   **Deep Learning** (ANN,RNN, LSTM)

Professional Experience

06/2022 – present
Pune

Data Scientist II, *Antuit*
- **Implemented** end to end ML Pipeline from the scratch for the Time series data by using models like ARIMA,SARIMA,LSTM,LGBM,SVMs and improved accuracy by 5%.
- Experience in creating mature Data science pipelines encompassing EDA,Featrure Engineering,Feature selection, Model selection, Model deployment.
- **Automated** different pipelines and day to day repetitive tasks by parameterized them using Python.
- **Achieved SLA** for customer deliverables by sharing data insights and weekly forecast analysis.
- **Written Robust and Scalable Python code** for implementing ML pipelines that will sustain for variety of data for different customer.
- Advance knowledge of **ML algorithms** such as Ensemble Models, Ada-boost, Gradient-Boosting, Kernel-KNN, Grid search,Random Forest, Decision Tree, SVM, Linear Regression, Ridge, Lasso, Pipeline, Feature Union, Cross Validation, KFold.

04/2019 – 06/2022
PUNE, India

Tech Analyst, *Cybage Software Pvt Ltd*
- Expertise in Data Analysis, statistics, programming ML/Al techniques and Big data technologies.
- Extensive Model building experience with Machine Learning/Al algorithms for Product developments.
- Involved in creating various **regression and classification algorithms** such as Linear Regression,Decision Trees,Naïve Baye's,Logistic Regression,Random forest,Ensemble techniques.
- Hands on experience on **feature engineering** techniques such as EDA,Missing value treatment, outlier detection, Normalization and Standarization.
- Involved in implementing **feature selection** techniques such as Correlation,Forward elimation,Backword elimination, univariate selection.
- Design and build ETL pipelines to automate ingestion of structured and unstructured data.
- Involved in **Data Pre-processing** Techniques such as data cleaning, visualizing the data, identifying outliers and making data ready for Machine Learning Techniques.
- Hands on experience on various **AWS cloud services** like Amazon Sagemaker,Amazon glue, Amazon S3 Amazon EC2,Amazon Cloudwatch,Amazon EMR.
- Experienced in running query - using Impala and used Bl tools to run ad-hoc queries directly on Hadoop.
- Created Regression, Classification and Clustering models for different data sets.

- Hands on experience on **hyperparameter tunning** techniques such as Gridsearch,RandomizedSearch,Hyperopt,Optuna
- Performing Residual analysis of the data with its Residual plot. Predicting and analyzing multi-collinearity in the models,
- Involved in various **NOSQL databases** like Hbase, Cassandra in implementing and integration.
- Experience in designing both time driven and data driven automated workflows using Oozie.
- Experience in Creating a Machine Learning Based Web Application Using **Django**

02/2018 – 09/2018 PUNE, India	**Process Executive,** *Spectrum Consultants (Nvidia)*

- Involved in creating various regression and classification algorithms by using various sklearn libraries such as Linear Regression,Decision Trees,Naïve Baye's
- To Pre-process the data of more than 1 Million Records and hundreds of Features.
- To **build and Train** a Machine Learning Model using Logistics Regression, Random Forest Classification, Xgboost Classification.
- To find the accuracy of the Model Prediction using Classification Reports, Confusion Matrix, AUC Score.
- Applied **Spark** advanced procedures like text analytics and processing using the in-memory processing.
- Responsible for Data Pre-Processing and Data Transformation using Python.
- Fluency with Python for data-science investigations using different libraries such as **Numpy, Scipy, Pandas, Scikit-Learn** in Python.
- To work cross-functionally to define problem statements, collect data, build analytical models and make recommendations.
- To lead and support various ad hoc projects, as needed, in support of Organization's Business strategy.

Projects

04/2019 – present	**To Predict Profitable Customers For Digital Marketing Product,** *Epsilon*

- Implemented various machine learning algorithm for the digital marketing product for various customers
- Handled data in PB and drawn useful insights from it.
- Recommended products with market-basket analysis techniques which lead to increase in customers by 40%.

02/2018 – 10/2018	**Segmentation Analysis For Hardware Products,** *Nvidia*

- Predict the rating of all customers in order to decide whether he/she is a potential customer to be considered by taking the history of the transactional data
- Involved in Data Pre-processing Techniques such as data cleaning, visualizing the data, identifying outliers and making data ready for Machine Learning Techniques.
- Machine Learning Algorithms Evaluation.
- Analyze the results of all algorithm outcomes and deciding the best one
- Achieved 30% increase in the sales of the product

01/2021 – 03/2021	**To predict student attrition for clearwater suites boise state university,** *Jigsaw Academy*

- Implemented various ML algorithm like Decision Trees,Linear regression for the real time data.
- Handovered the analysis result and recommendation to university to reduce the attrition
- Achieved decrease in attrition by 50%

Certificates

- Data Science Specialization by IBM
- Full Stack Data Scienctist by Jigsaw Academy
- Machine Learning and Deep Learning masters by Ineuron

Education

06/2013 – 08/2017 Kolhapur, India	**B.E Mechanical,** *KIT's COEK,Shivaji Univeristy*