Zaid Ahmad

Data Scientist - US Healthcare

Location Preference: Delhi/NCR
Email: syedzaidahmad99@gmail.com
Mobile: +91-8826210811

Business Analyst
Exl Service Pvt Ltd, Noida

Personal Details

Date of Birth :05th July 1993
Address: 99 D Zakir Nagar, Delhi
Languages Known: English and Hindi
Hobbies: Cooking, Chess

EDUCATION

Master of Computer Application
Jamia Millia Islamia
07/2016 - 06/2019, Delhi

Bachelor of Science
Allahabad University
07/2012 - 06/2016, Allahabad

Profile Summary

- 2+ years of strong experience in Data Science, Data mining with large data sets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data modeling, Data Visualization, Web Development,and Apache Spark including Big Data technologies like Hadoop, Hive, Pyspark
- Strong experience in core python along with object-oriented programming
- Excellent knowledge of Machine Learning with Python, Linux/Ubuntu, Relational databases, and shell scripting. Deep understanding & exposure of Big Data Eco – system
- Expertise in managing entire data science project life cycle and actively involved in all the phases of project life cycle including data acquisition, data cleaning, data engineering, data miningalgorithm, features scaling, features engineering, statistical modeling (decision trees, regression models, neural networks, SVM, clustering), dimensionality reduction using Principal Component Analysis and Factor Analysis, testing and validation using ROC/AUC plot, K- fold cross validation and data visualization.
- Extensively used SQL, CV2, H2O, NumPy, Pandas, Scikit-learn, Spark, Jupyter-Notebook, Hive for Data Analysis and Model building.
- Strong experience and knowledge in Data Visualization with Python creating: Line and scatter plots, Bar Charts, Histograms, Pie chart, Box plots, subplots etc.
- Experienced with end-to-end web development using Flask, MySQL and HTML/CSS.
- Effective communicator with leaders and team members; briefing team on project and program statuses and results;

Technical Skills

- Python, Object Oriented Programming, C/C++, Data Structure, Algorithm
- MySQL, Hive, PySpark
- Scikit-learn, NumPy, Pandas, Sea-Born, Matplotlib, GIT, JSON
- Flask, HTML, CSS, OpenCv
- Shell Scripting, Linux/Ubuntu
- Hadoop, Big-Data
- Others: Adequate Hardware Knowledge

PROJECTS

Since June 2019 – Present, Exl Services Noida

Algorithm to predict if a person will get eye surgery

Developed the model based on the member's historical data (Claims, Lab, and Pharmacy Data). Feature extraction has been done through HQL (Hive Database). Built XGBOOST, LightGBM classifiers in python using Scikit-Learn to predict which patients might be get an eye surgery. Applied cross validation (CV=5) on the training data and Randomized search CV to select the best parameters of the model. Determined model performance by comparing recall and precision on test data to choose the best Machine Learning Algorithm.

Fertility test prediction	Performed explanatory Data Analysis and applied extensive feature engineering, feature selection on female's data. Data preprocessing and model development has been done on python (using Scikit-Learn). Applied cross validation (CV=5) on the training data and Randomized search CV for hyperparameter tuning. The model has been developed using Light GBM.
Optical Character Recognition (Web development)	Extract patient details from US health care form which is in jpg format. Applied various image processing technique such Gaussian filter, Threshold filter etc. using CV2, then extract text from images. Stored it to MySQL databases. Created web pages using Flask, HTML/CSS for revalidation whether image extraction has been done successfully or not. If any value missed out/unable to extract, that value can be fill up manually using that web page.
Dashboard	Created dashboard for Data Science project applicable for various project for data visualization. Design Algorithm using python (with OOPs) along with its various libraries such as NumPy, Pandas, Matplotlib etc. Voila, PySpark, hive, Hadoop, and shell scripting. This dashboard runs daily using crontab in putty and data visualization displays on browser to any system via any free port using url.

Awards and Certification

- Award for developing JSON QC script using python, PySpark, and Hive which reduces human efforts and time. It is generalized QC script applicable almost on all JSON.
- Problem Solving in algorithm achieve 1736 (5 star) on Hacker Rank.
- Client always recognized the work through appreciation mails