|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| my_img.jpg    **Vibhanshu Swaroop Srivastava**    +91 9871394498  vibhusw@yahoo.com  Result-oriented professional, targeting mid/senior-level assignments in **Data Science** with a leading organization of repute, preferably in **Pune, Bengaluru, Delhi/NCR**  Core Competencies   |  | | --- | | * Data Analysis & Algorithms | | * Analytics Solutions Implementation | | * Machine Learning models | | * Dataset Management | | * Statistical Analysis Program Execution | | * Python | | * Cross-functional Coordination | | * NLP, NLTK , Bag of words | | * PCA & Boosting Techniques | | * Logistics & Linear Regression | | * Decision Tree |   Soft Skills    Logical  Analytical  Team Leader  Change Agent  Collaborator  Motivator  Communicator  Achievements   * Collaborated cross-functionally and implemented analytics and visualization components for device data analysis platform to predict hardware * Assisted majorly in the optimization of factors for increase in model performance * Coordinated for implementation of models(like Random forest, Regression),a tool specialized in identifying crashed processes and help restart the process automatically   Academic Details   * B.Tech. (Electronics & Communication) from IMS engineering College, Ghaziabad in 2015   Languages Known  English  Hindi  Personal Details  **Date of Birth:**12th Sep1991 **Address**: Blue Ridge, Pune | Profile Summary   * A result-oriented professional with over 3+ years of experience in **Data Science** * Identified, developed and implemented appropriate statistical techniques, algorithms, and **ML Models to create new, scalable solutions that resolved business challenges** * Experience in creating **Data Science Pipelines** encompassing Data Standardization, Feature Extraction, Model Validation and optimization using Python * In-depth expertise in **Text & sentiment analysis using NLP, NLTK libraries**. **Regression, Classification, Clustering and Dimensionality Reduction algorithms** * Hands-on experience in **building/validating predictive models** using techniques like **Tree based algorithms- Random Forest** using **Python.** * Used the **latest data and visualization tools** and software to provide the business with insights on customers and make recommendations for improvement * Gained experience in the definition, development and evolution of **data models, tools and capabilities** * Evaluated the effectiveness and accuracy of **new data sources** and provided timely communications on significant issues and developments * Developed **customized data models** and algorithms to apply on data sets * Displayed skills in developing and training other Data Scientists * Developed **predictive models**& machine learning algorithms   Technical Knowledge   * Experience in creating models using algorithms Regression, Decision Trees, Random-Forest, Boosting techniques * Visualization experience with ggplot, seaborn and matplotlib in Python * Expertise in Python for data-science investigations using different libraries such as NumPy, SciPy, Pandas, Scikit-Learn in Python * Text & sentiment analysis using NLP, nltk libraries * Knowledge of TF-IDF, Bag of words & Word 2 Vec techniques.   Work Experience  **Accenture Pune**  **Data Scientist**  **May’17 till present**  **Key Result Areas:**   * Collaborating with multi-disciplinary teams to understand business requirements for the products and projects * Analysing statistical data; importing data from multiple sources, joining datasets, cleaning & preprocessing, data wrangling and visualizations * Identifying and implementing statistical models including linear models, multivariate analysis, stochastic models, sampling, optimization and time series analysis * Applying machine learning and deployment of models * Querying data from various database and creating data sets that can then be analysed to provide key insights * Manipulating data to fit the requirements of the analysis * Constructing data sets by understanding the purpose of the data and the story it is going to tell * Ensuring compliance with coding standards & internal practices * Diagnosing bugs and data flow issues - reviewing old code to ensure that data is flowing correctly and fixing when they arise   **Project 1:** KEES (State of Kansas)  **Period**: Mar’19-Present  **Problem Statement 1**: Check eligibility of person for medical benefits.  **Problem Statement 2**: Prediction of amount allotted to citizen based on Treasury offset program.  **Algorithms used:** Logistic & Linear Regression, Decision Tree, Random forest.  **Role**:   * Creating Feature engineering pipelines encompassing Data pre-processing techniques, Data standardization, and Feature extraction * Developing Regression, Clustering & Classification models for different data sets * Predicting and analyzing multi-co linearity in the models * Managing bias-variance in Linear Regression Models using Regularization techniques * Using ggplot, Seaborn and Matplotlib for visualization * Performing analysis of ROC & AUC curve for the binary Classification data * Reducing dimensionality from PCA & Scree plots * Tuning hyper parameter using Grid search CV * Using Boosting techniques to enhance the performance & stability of the model * Engaged in Model monitoring & retraining   **Project 2:** Virgin Media UK  **Period**: May’18-Feb’19  **Problem Statement**: Sentiment Analysis of various products based on customer’s reviews  **Algorithms/Techniques used:** NLP/NLTK, Bag of words, Stop words, TF-IDF, Word 2 Vec, Stemming/Lemmatization, Logistic Regression, Decision Tree, Random forest.  **Role**:   * Developing Feature engineering pipelines encompassing Data pre-processing techniques, Data standardization, and Feature extraction * Creating Classification models for different data sets * Tokenization, data cleaning by removing stop words, punctuations. * Generating root words from text using Stemming & Lemmatization techniques * Analyzing frequency of different words with help of Bag of Words, n-gram models. * Numerical representation of words using TF-IDF ,Word 2 vector techniques * Using Boosting techniques to enhance the performance & stability of the model * Engaged in Model monitoring & retraining |