Srijan Sahay

Data Scientist

Experienced Data Science professional with sound programming and theoretical skills. An eager and quick learner who can understand and implement new technologies and models rapidly.

Experience

Aug-2019-Present

Data Scientist
Motifworks India Pvt. Ltd.,

- Delivering end to end Microsoft Azure solution
- Using machine learning techniques to select features, build, and optimize classifiers
- Performing data mining by utilizing latest state-of-the-art processes
- Dealing on company's behalf with third party sources of information as and when required
- Undertaking innovative methods of data collection building information that is relevant to create analytic systems
- Processing, sorting, and double-checking data integrity for proper data analysis
- Doing additional analysis to present results in a clear and systematic manner.

Dec-2016 - Aug-2019

Senior System Engineer
Infosys Ltd.,

- Using machine learning techniques to select features, build, and optimize classifiers
- Processing, sorting, and double-checking data integrity for proper data analysis
- Creating courses on Data Science and Machine Learning
- Creating internal certification for Infosys employees
- Creating tools and solution for Infosys internal use

Hackathon Experience

Oct-2018 - Dec-2018

Yes Bank Datathon
Finished in the top 21 teams out of 1700 teams in a three month long hackathon organized by Yes Bank Ltd, one of the largest private sector banks in India, in December 2018. As part of this hackathon:
- Used transaction table which was 50 GB in size to classify transactions as outflows and aggregated them for each customer at a monthly level.
- Developed features like lags, cumulative outflows, number of transactions for each customer for each month and joined it with demographic and service request tables
- Trained a Random Forest Regressor on the developed features that predicted the monthly outflows for 180,000 customers. The Symmetric Mean Average Percentage Error(SMApe) stood at 38% for the test period.

Certificates

Jan-2020
Microsoft Certified: Azure Data Scientist Associate

Jun-2019
MTA: Introduction to Programming Using Python - Certified 2019

Oct-2018
Certificate for successfully completing Capgemini Tech Challenge 2018

Education

Aug-2012 - Jul-2016

Gandhi Institute for Technological Advancement
Computer Science & Engineering, B.Tech

Apr-2009 - May-2011

Chinmaya Vidyalaya
12th CBSE

May-2009

Chinmaya Vidyalaya
10th CBSE

Personal Info

Address
Pune

Phone
+91 7205255668

E-mail
srijan.bhushan07@gmail.com

Date of birth
07-July-1993

GitHub
https://github.com/srijansahay

LinkedIn
https://www.linkedin.com/in/srijansahay

Kaggle
https://www.kaggle.com/srijansahay

Medium
https://medium.com/@srijansahay

Skills

Python

numpy,pandas,matplotlib,sklearn, nltk,opencv

Data Analysis

Machine Learning

Naive Bayes,SVM, KNN, Tree-based Methods, Clustering

Deep Learning

Cnn, object detection, object classification

Keras

Azure Services

Achievements

- Won "Client Recognition" award for performance in Infosys
- Insta award by Infosys for meeting time commitment for a client

Miscellaneous Projects

Performed analysis on Novel Corona Virus to find the trend of it spread and death caused by it.

Developed a Deep learning based image classifier to accurately classify hand-written digit for mnist dataset.

Developed a regression model to predict the house price on KC_House dataset.

Developed a classifier to predict the permeability of the rock sample.

Languages

Hindi - Native Language

English – Professional Efficiency

Aug-2019 - Present

Data Scientist
Motifworks India Pvt. Ltd.,

Outlier Detection

Optimization of Fuel invoice validation for a US-based retailing, fuels marketing, and commercial real group of companies. The model was able to find the inappropriate ancillary charges in the fuel invoice and flag them to the freight manager reducing the manual effort.

Responsibilities:
- Collecting data from different sources into Azure Data Lake.
- Cleaning and pre-processing the data.
- Applying and researching different clustering algorithm to find the outlier in ancillary charges.
- Deploying this model as web services and connecting it with PowerBI for better report generation.
- Flagging ancillary charges with outlier in the report.

Sentiment Analysis

Opinions are central to all most all human activities because they are the key influencers of our behavior. Created a model that took the product reviews of an organization and gives the customer's sentiment about that product.

Responsibilities:
- Collecting data from different sources into Azure Data Lake.
- Pre-processing of review data like Sentiment sentences extraction and POS tagging.
- Negation phrases identification and Sentiment score computation for sentiment tokens.
- Research on types of classification algorithm that can be used in this scenario.
- Deploying this model as web services and connecting it with PowerBI for better report generation.

Customer Segmentation

Created a clustering model to find various patterns in customer purchase behaviour to identify different categories of customers. Targeting each category of customers in different ways to increase the business.

Responsibilities:
- Collecting and cleaning the data from different sources.
- Finding useful insights on processed data.
- Finding loyal and wandering customers.
- Finding ways to retain loyal customers and converting wandering customers into loyal customers.
- Research on different types of clustering algorithm that can be used.

Dec-2016 - Aug-2019

Senior System Engineer
Infosys Ltd.,

Car Price Prediction

Optimization of Second-hand car Sales for a US-based auto retailer. The model was able to predict the price of a given car at which it could be sold easily and when sold to a customer, company was able to make a great profit.

Responsibilities:
- Identified the important driver variables and correlation between them which can help the company get into a new market.
- Research on the types of regression algorithm that could be used.
- Tuning hyperparameters of the model.
- Boosting to minimize the error value by avoiding overfitting.

Churn Prediction

Developed a Machine Learning model to predict the probable exiting customer for a European telecom service provider. The model was able to predict the churn rate and identify the customer who may leave the network.

Responsibilities:
- Collecting and cleaning the data from different sources.
- Identifying the key features and variable for the customer churn
- Research on types of classification algorithm that can be used in this scenario
- Creating a Machine Learning model which when given the features of a customer was effective to predict the probability of churn.

Face Detection And Recognition

Developed a demo using Python's Dlib library for a fast-food chain that kept track of the count of unique customers at service counters. The model detected faces, generated and compared embeddings to come up with the count of unique customers in the queue.

Responsibilities:
- Research on the types of algorithms that could be used.
- Prepare training data to retrain the weights of various different models, and evaluate the performance of all different models used.
- Tuning hyperparameters of the models using common techniques like GridSearch
- Benchmarking the performance of NN based model against conventional ones.

Automated Essay Scoring

Developed an essay scoring algorithm and implemented it to create a machine learning model. The model is used to score essays in Infosys training.

Responsibilities:
- Created different corpus for this scenario.
- Identified and created features for predicting the essay score.
- Designed an essay scoring algorithm for Infosys.
- Reduced manual effort for scoring the essay.

Automated Machine Learning Certification

Developed an automated machine learning certification for Infosys internal certification.

Responsibilities:
- Managed Git for creating the assessment.
- Developed 8 machine learning problem with solutions and test cases in R and Python.
- Developed 100+ objective questions.
- Populated docker to deploy docker.
- Established an automated machine learning certification for Infosys Internal Assessment.