

Sathish Kumar

Data Scientist - HDB

Singapore

- Email me on Indeed: <http://www.indeed.com/r/Sathish-Kumar/1e39b2df4e494f4f>

- Curious, enthusiastic and self-motivated analytics professional with 4.1 years of experience in understanding business problems, data acquisition, data pipelining, data pre-processing, data transformation, building machine learning models, data visualization, Optimizations.
- Having good experience in various probability, statistical and machine learning algorithms (unsupervised, supervised, predictive, classification, reinforcement, regression and clustering algorithms), Deep learning, NLP and image processing.
- Worked on analyzing large data sets using statistical and analytical knowledge to generate key business insights, understand current data trends and forecast outcomes in order to make informed business decisions.
- A deep hands-on experience in Relational/NoSQL databases.
- Strong team player, practice continuous learning & strive to achieve higher standards in future roles.

Work Experience

Data Scientist

HDB - Singapore

October 2018 to Present

- Handled huge financial data using Data pipe lining techniques and transformed raw into statistical features for applying machine learning / Deep learning algorithms to derive useful insights.
- Developed insightful and user friendly dashboards using Tableau and Matplotlib.
- Worked closely with cross functional financial teams, to understand the pain points where machine learning can ease the works.
- Involved in transforming business problems into statistical problem statements and choose most efficient machine learning models by considering the various KPIs.
- Developed & deployed various predictive, regression and classification models.

Data scientist

Sopra Steria India Ltd - Chennai, Tamil Nadu

August 2015 to October 2018

Developed end to end machine learning model and worked in deployment of ML models in production environment.

- Involved in discussion with business people to understand the requirements and decided which category of data needed for training and testing the ML model.
- Involved in collecting data from different cross functional businesses (live data) and stored in appropriate data stores (Relation/NoSQL databases, different file formats) by using data pipelining methods.
- Transformed raw data into statistical features and applied most efficient Machine learning models (predictive, recommendation, regression, classification, and forecasting and ensemble models) to achieve the high accuracy model.

- Determined important KPIs needed for the ML models and achieved it with high accuracy by conducting various metrics techniques (Confusion matrix, Classification accuracy, ROC, precision, etc.,).
- Did A/B and hypothesis testing for ML models to ensure the accuracy and efficiency of the model.
- Developed dashboards using Tableau.
- Built various in-house products (predictive, recommendation and NLP models) which reduced the man power and cost for the client.
- Built an image processing deep learning model which reduce the insurance claim time for the customer and eased the document process.
- Received GEM and ACE awards for successful implementation of ML models which increased the number customer and revenue for the client.

Education

Bachelors in Computer Science and Engineering

Mahendra Engineering College, Anna University

2011 to 2015

Skills

Python (4 years), R (3 years), Machine Learning (4 years), Deep learning (4 years), Tableau (4 years), NLP (2 years), Image processing (2 years)

Links

<http://www.linkedin.com/in/sathish-datascientist>

Additional Information

TECHNICAL SKILLS

Programming Languages: Python, R, and SAS.

Database: Postgre, mongo DB, Cassandra, SQLite.

IDE: Spyder, IntelliJ Idea, Pycharm, Jupyter, Anaconda.

Data Visualization: Tableau, Matplotlib, ggplot.

Algorithms: Regressions, Clustering, Dimensionality reduction, Forecasting, Decision tree, Random forest, KNN, K-means, SVM, Naïve Bayes, CNN, ANN, RNN, Deep Boltzmann Machine, XG Boosting, Regularization, reinforcement algorithms.

Libraries: Scikit learn, Pandas, Mlpy, NumPy, SciPy, Matplotlib, tensorflow, NLTK, Keras, Theano, Pytourtch, Tkinter, spacy.

Big Data: Hadoop, Spark, Sqoop, Flume, Apache storm, pig, hive, map reduce.