Hemanth Ravikumar
Data Scientist - Avid Data Science Services

Arlington, TX - Email me on Indeed: indeed.com/r/Hemanth-Ravikumar/90dbbcd0b9e97d9d

• A data science professional with over 6 years of hands on experience in Data Science with clear understanding of various concepts in machine learning, Text mining & statistical modeling techniques; and hands on experience on implementing them to solve complex business problems.
• Have good knowledge on various descriptive statistics and inferential statistics like Normal Distribution, t Distribution, Chi-square, F- Distribution and hypothesis testing.
• Deep understanding of the techniques in Machine learning and Statistical algorithms in forecasting, classification and optimization.
• Strong academic expertise in Natural Language Processing, Speech Processing and advanced machine learning.
• Have a good work experience on
◦ Statistical and decision modeling techniques like linear, logistic regressions and time series analysis.
◦ Various supervised machine learning algorithms like Decision Trees, Support Vector Machines, Random Forests and K-NN.
◦ Clustering techniques such as K-Means, Hierarchical clustering.
◦ Deep learning and Neural Nets.
• Hands on work experience with Big Data eco systems components.
Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist
Avid Data Science Services   -   McKinney, TX - January 2017 to Present

Project: Fuel Theft Analysis
Perform fuel analytics to visualize the fuel consumption per asset/vehicle (movable and non-movable) with the help of various sensors attached to them. The objective was to identify the rate of fuel consumption per equipment and account for the instances of sudden decrease in fuel levels.

Key Contributions:
• Understand the type of sensors deployed such as Float and Capacitive sensors
• Define the target objects based on the business needs such as rate of fuel usage per equipment and sudden decrease in fuel level
• Data Preprocessing
• Exploratory data analysis
• Address challenges such as faulty sensor readings and meter readings
• Algorithm Development - LOESS method/ Anomaly detection method
• Aggregate massive volumes of records
• Identify fuel fill and fuel dipping instances

General Doctor Advise etc
Bridgei2i Analytics Solution Private Limited - November 2013 to November 2014

Project: Text Categorization Nov 2013 - Nov 2014

The client follows a ticketing system for all the telephonic calls received across all the departments. Calls to the provider can be for New Appointment, Cancellation, Lab Queries, Medical Refills, Insurance Related, General Doctor Advise etc. The Tickets have the details of Summary of the call and description of the calls written by various staff members with no standard text guidelines. The challenge is, based on the Text in the Summary and Description of the call, the ticket is to be classified to Appropriate Category (out of 5 Categories) and Subcategories (Out of 20 Sub Categories).

Approach:
• Cleaning the data using stemming and lemmatization methods
• Extracting the key words from summary and call description
• Keywords matching with target classification variables
• If it matches tag them into respective category and then in to sub category based on the key words

Key Skills: Text Mining, Naïve classifier, KNN classification, Support vector Machines

Data Scientist
Bridgei2i Analytics Solution Private Limited   -   Bangalore, Karnataka - November 2011 to November 2014

Data Scientist
Bridgei2i Analytics Solution Private Limited - January 2012 to October 2012

Have used data ingestion tool, PySpark and Sqoop for web log data loading, and processing.

Key Contributions:
• Loading data from SHOP CA website to HDFS using flume.
• Doing the preprocessing of log files using Apache log format.
• Analyze the log files using PySpark and store the results into local file using SQOOP.
• I was involved in preparing understanding document of the Business scenarios.
• Prepared complex queries for the adhoc reporting and designed dashboards in Tableau.

EDUCATION

Master of Science in Electrical Engineering
University of Texas at Arlington   -   Arlington, TX
December 2016

B.S. in Electronics and Instrumentation Engineering
Amrita School of Engineering   -   Bangalore, Karnataka
May 2012

SKILLS

Sqoop (Less than 1 year), HDFS (Less than 1 year), HADOOP DISTRIBUTED FILE SYSTEM (Less than 1 year), APACHE HADOOP SQOOP (Less than 1 year), APACHE HADOOP HDFS (Less than 1 year)

ADDITIONAL INFORMATION

Skills

Analytical Tools R Studio, IPython, MATLAB, Git, Wireshark, IDA Pro, OpenCV
Big Data Eco Systems HDFS, Spark, MapReduce, Hive, Pig, Oozie, Sqoop, Flume, Mahout,

Programming/Scripting Languages C/C++, Assembly, Java Script, Go Lang, Shell Coding, R, Python, SQL, Pig, Hive
Databases Oracle, MySQL, Mongo DB, HBase

Data Science Skills

Statistical Modeling Regression Models, Time Series Analysis, Bayesian Methods
Regression Techniques Logistic Regression, Linear Regression, Multi-Linear Regression
Machine Learning
Decision Trees, Ensemble Methods, Neural Nets, ANN & RNN, PCA, Rules, Clustering, SVM, K-NN, Deep Learning, Evolutionary Search Methods

Text Mining
Language Modeling, Text Indexing, Web Crawling, NLP, Sentiment Analysis, Text Categorization & Summarization, Social Network Analysis

Relevant Technologies Data Science, Machine Learning, NLP, IoT, Image Processing, Speech Processing, Computer Vision, AI, Information Security

Key Skills: Cloudera CDH 4, HDFS, flume, PySpark, SQOOP, Tableau, and Linux

Project: Dynamic pricing model Nov 2011 - Dec 2012
Developed a dynamic pricing model using predictive analytics techniques for a leading logistics client. Enhanced the algorithm to scale it to the orders with a new pair of source and destinations using ensemble algorithms.
Key Contributions:
• Understanding the business context to understanding the data and to develop the conceptual model.
• Data Preparation and pre-processing
• Model building and Evaluation
• Client interactions

Key Challenges:
• Outliers and noisy data
• No standard price charged to the shipments in the same route
• Multiple Zip codes which leads to too many variables while building the model
• Inconsistent date formats
• Duplicate orders
• Multiple data resources
• Extreme non-linearity in the data
• Arrive at a metric to measure the confidence of the predictions generated from the Ensemble Algorithms
• Technical integration of machine learning models in R and user interface in Python.

Key Skills: Data Preprocessing, deep learning, time series, decision trees (CART), random forest, R, python

Bridgei2i Analytics Solution Private Limited Bangalore, India
Junior Data Scientist Mar 2011 - Oct 2011

Project: Prediction of A & B Link failures
Identify pre cursors to A & B Link failures; predict the likelihood of A & B Link failure in next 3 days.

Approach:
• Period selection - such as defines the target event and defines observations window & performance window.
• Data Preparation - identification of potential predictors, consolidated view of device at daily level.
• Model building - build the A & B link failure propensity models by implementing decision trees/logistic regression algorithms.
• Iterate and refine the models for better results to meet business requirements.

Key Contributions:
• Finding the key predictors for A & B Link failures
• Analyzing the patterns of Network performance
• Build the predictive model CHAID Model to identify the pre-cursors of A & B link failures.
• Results Analysis.

Key Skills: Decision trees, logistic regression

Academic Projects and Internships

• Smart traffic decongestor May 2011 - May 2012
Worked with Bangalore city Traffic police to control traffic menace in Bangalore, India. Active RFID tags and RFID readers were used to estimate traffic density. A five layer Fuzzy neural network with gradient descent as a training algorithm and genetic algorithm were used to build artificial neural network to control the traffic signal. Simulations were carried out on matlab and observed a reduction in mean delay and stoppage time of each vehicle by over 80%.

• Machine learning approach to detect ransomware(case study) Nov 2015 - Apr 2016
Cuckoo sandbox-ing to analyze and extract over 30,000 features and feature scaling to top 100 most relevant features on a dataset consisting of ransomware and goodware applications. Used logistic regression to detect if the application is goodware or ransomware with an accuracy of over 96% and analyzed its ability to detect new ransomware families and suggested improvements.

• Digital Video Coding Aug 2015 - Jan 2016
Worked at Multimedia processing lab at UTA with the co-inventor of discrete cosine transform (Dr. K.R. Rao), on techniques like DCT, slant transform, wavelet coding, JPEG-LS, JPEG-lossless and multimedia compression algorithms and implemented them on Matlab.

• Digital Image Processing Feb 2015 - Dec 2016
Worked at Multimedia processing lab at UTA with Dr. K.R. Rao, on various image processing techniques like quantization, histogram equalization, inverse Gaussian filtering, edge detection, geometric filtering, inverse and Weiner filtering and implemented them on Matlab.

• Recognizing handwritten digits Sep 2016 - Dec 2016
Used all-vs-one logistic regression and neural networks to recognize handwritten digits for a given dataset and implemented it on python and matlab. Obtained 94.9% accuracy on training set for all-vs-one logistic regression, 97.3% accuracy for feedforward propagation and 99.3% when using backpropagation.