Vinay Kumar R

vinayvinkumar@gmail.com; 9036172033
677 4th Cross Mahalakshmi Layout, Bangalore-560086
linkedin.com/in/vinkumarr; github.com/vinayvinkumar

SUMMARY

Data Science enthusiast with 6 years of industry experience and special interests in Machine Learning techniques, Big Data Analytics, Data Visualization, Statistical Modeling and Predictive Analysis. Skilled in statistical analysis of A/B testing new features. Passionate about achieving a challenging position that allows contribution of my analytical skills to an organization's success and personal growth.

EDUCATION

Master of Science in Data Science, Indiana University Bloomington, USA **MAY 2017**
Bachelor of Engineering in Computer Science, Visvesvaraya Technological University, India **JUNE 2010**

TECHNICAL SKILLS

- **Programming**: Python {scikit-learn, pandas, numPy, matplotlib, sciPy, iPython notebook}, R {ggplot2, dplyr, data table}.
- **Machine Learning**: Regression, Decision Trees, Logistic Regression, Gaussian Naïve Bayes, Neural Networks, k-NN, Clustering, Ensemble Learning, Boosting (AdaBoost, Gradient Boost, XGBoost), Support Vector Machines, Cross Validation, Principal Component Analysis, Natural Language Processing.
- **Big Data:** Spark, Hadoop MapReduce, Hive, AWS.
- **Database**: SQL, SQLite, PostgreSQL, MongoDB, SAP Hana.
- **Statistical Methods**: Hypothesis testing and Confidence Intervals, Regression models, Probability.
- **Miscellaneous**: Git, A/B testing, Microsoft Excel, Selenium

DATA SCIENCE PROJECTS

- **Predictive Analytics**
  - '**Credit Modeling**': Built a Machine Learning model to accurately predict the probability of borrower paying off loan on time. Performed Data cleaning, Data preparation, handled missing values, converted categorical columns into numbers before modeling the data using Logistic Regression and Random Forest classifier. Used K-Fold cross validation technique to overcome the problem of over-fitting.
  - '**Sentiment Analysis**': Built a Classification model using Gaussian Naïve Bayes, Decision Trees to 'predict whether a customer review is negative or positive, based on the text alone'. The algorithms were implemented from scratch without using any per-built library functions.
  - Participated in **Kaggle** '**Titanic: Machine Learning from Disaster**' learning competition to predict 'which passengers survived the sinking of Titanic' and ranked within top 21% of the competitors.
  - '**Uber challenge**': Predicting 'Which driver sign-ups are most likely to start driving'. Performed Data Cleaning, Data Visualization, Feature Engineering, and Dimensionality Reduction before implementing Machine Learning model Random Forest Classifier and Gradient Boosted Trees.
  - '**Iris Species Classification**': Built a predictive model using Neural Networks to classify the species of Iris flowers (Virginia or Versicolor) using features like sepal length, sepal width, petal length and petal width. Back-propagation algorithm was used to update the parameters for a three layered Neural Network. Benchmarking was done using roc_auc_score to check the performance of the Neural Network.
- **Data Science Programming**
  - **Natural Language Processing**: NLP using 'Bag of Words' technique to transform the written text into a numerical representation, and to use that representation to make predictions. The project made use of "Hacker News" article headlines to predict a number of up votes the articles received. Since up votes are an indicator of popularity, exploration was done to discover which types of articles tend to be the most popular.

- Coded Linear Regression, Logistic Regression, Naïve Bayes, Decision Trees, Neural Networks, Expectation Maximization and many other Machine Learning algorithms in Python and R.
- **'Bike rentals prediction'**: Project demonstrating predictions of bike rentals using Linear Regression, Decision trees, and Random Forest. High Skewness on the distribution of the counts of bike rented affected all three Machine Learning models due to which the MSE turned out to be very high.

- **Data Analysis**
  - '**Pixar's movie review**': Analysis of how Pixar movies are rated by rotten tomatoes, Metacritic score, and IMDB rating. Conclusion – Rotten Tomatoes gives Pixar movies higher rating than any other critics.
  - '**Star Wars Survey**': Analysis of 'Star Wars' Survey data which was collected by 'FiveThirtyEight' when a new Star Wars movie was about to release. Conclusion – Star Wars Episode 5 had received best rankings and hence was people's favorite.

PROFESSIONAL EXPERIENCE

SAP, Bangalore, India
Senior Software Engineer **May 2014 – July 2016**

- Led Enterprise Mobility Management project related activities of the whole team in terms of Trend analytics, Device usage analytics and designing custom reports resulting in making better decision making on the product road map.
- Interacted with customers by gathering requirements, designing, epics and story point's creation serving a smooth project life-cycle.
- Led 'Sentiment Analysis' project using SAP HANA in order to predict the customer churn by making use of historical data.

SAP, Bangalore, India
Software Engineer **Oct 2010 – May 2014**

- Contributed in developing more than 200 highly reliable automated scripts for SAP banking analytics scenarios helping the team in quick product delivery and customer satisfaction (***Won Spot Award for quick delivery***).
- Cultivated frequent interactions with stake holders in understanding new requirements and preparation of weekly status reports.
- Performed analytics on banking transactions to help the team's product delivery by properly identifying prioritization of automation backlogs (***Won Spot Award for taking right initiatives***).

CERTIFICATIONS

- Data Analyst specialization track – Dataquest.io. (License: **YII1KWGILWO1YGPQM39D**)
- Data Scientist specialization track – Dataquest.io. (License: **T890TAPHDLUGM37TDVRG)**
- Scrum Master Certified (License: **540785**).
- Certified in SAP
  - Text Analytics: HANA Platform.
  - Introduction to Software Development on SAP HANA.

AWARDS/RECOGNITIONS

- Spot award for quick delivery of automation framework and scripts that helped customers run their end-to-end scenarios, ***SAP Labs India Pvt. Ltd.***
- Spot award for performing a deep analysis study on various Banking scenarios that helped to prioritize the automation backlogs, ***SAP Labs India Pvt. Ltd.***
- "Run Faster" award for automating the maximum number of scripts during a certain year, SAP Labs India Pvt. Ltd.
- "Run Stronger" award for time-to-time delivery of the bug free product to the customers, SAP Labs India Pvt. Ltd.
- Active participant of an intra-SAP hackathon competition "What the Hack", SAP Labs India Pvt. Ltd.

LEADERSHIP/INVOLVEMENT

- Collaborated to present a paper titled 'Image and Video retrieval' at a national level paper presentation KSIR 2009.
- Captain of the college table tennis team and won three doubles titles back to back in intra-college tournament, *Atria Institute of Technology*.