Vishal Babu *Nov10th, 1989*
Sr. data scientist @ blackstraw.ai
M.Tech (CSE), National Institute of Technology Karnataka, Surathkal
vishalbabu.in@gmail.com https://www.linkedin.com/in/vishalchak/
Singapore: +65-84555722 https://github.com/VishalChak
India: +91-8802901311

Profile Summary	5+ years of experience in Data Science & Development. Expertise in machine learning (ML), deep learning (DL), natural language processing (NLP), Computer Vision, NLU, Auto_ml (TPOT, autokeras) data science, artificial intelligence (AI), and software engineering. Worked extensively on python, pyspark, sparkML, Neural Network using Tensorflow, keras, and pytorch, Text mining using NLTK and spacy, NLP using Stanford NLP, AutoML using TPOT, Autokeras, and Autosklearn, NLU (RASA-NLU) etc.
Education	**M.Tech (CSE-IS)** - 6.3 - (2013-2015) - National Institute of Technology Karnataka, Surathkal (NITK). **B.Tech (CSE-IT)** - 69.92 - (2007- 2011) - United College of Engineering & Research, Greater Noida (UCER)
Technical Skills	**Programming:** Python, Java, PySpark, R and C. **Data Science:** Machine Learning, Deep Learning, AI, neural network, sk-learn, spark, spark ML, pySpark, NLTK, Sql, TPOT, AutoKeras, Stanford NLP, Keras, Tensorflow, computer vision, PyTorch, BERT, etc. **Databases:** MySQL, MongoDB & PostgreSQL. **Simulators:** Wire shark, ns2, ns3, BookSim, NoCSim, GpNoCSim **Other:** Git, Perforce, Svn, Tfs, EBit, Docker, web technologies.
Professional Experience	**Blackstraw.ai** Aug 2019 – Present At blackstraw I am working as a Sr. Data Scientist and worked on many interesting projects like real-time object detection classification as tracking (ODCT), Anomaly detection in manufacturing, and autonomous navigation and mandarin to English machine translation etc. **Equinix.** Sep 2018 – Aug 2019 Worked as a Data Scientist. Responsible for end to end data science project delivery from POC to project delivery. At Equinix, I am responsible to identify ML use cases, do POC on it, convert POC into project or services and integrate it with the existing product. **Larsen & Toubro Infotech (LTI).** July 2016 – Sep 2018 Worked as a Machine Learning Engineer. Responsible for doing RnD on new technology and end to end delivery from POC to project delivery. At LTI I am responsible for doing R&D on new technology, convert data into information, understand problem statement from client, requirement gathering, design and implement prediction models, and deliver it as a product or as a service to the client.

	AutoRABIT Inc. July 2015 – July 2016 Worked as a Backend Developer and developed modules, API calls by using Java, Javascript, and Jquery. Here I have written API calls, designed services as per requirement and research-development-integration of the module from scratch to end. During this period EZ check-in was done, user-based auto-commit done, development-integration of GIT, SVN, TFS, Perforce and written a wrapper for GIT hosted TFS etc.
Professional Projects	**Object detection classification and tracking (ODCT):** In this use case, we have created a real-time object detection classification and tracking system for BFL. This solution to identify treat by a camera and generate an alert. For object detection, we have done transfer learning over YOLO (You only look once) in darknet and trained a Deep-SORT (Simple Online and Real-time Tracking) in PyTorch on mars and Nvidia AI city dataset. **(YOLO, Deepsort, PyTorch, market 1501, darknet).** **Anomaly detection in manufacturing:** We developed a pipe-lined solution for detecting anomalies in manufacturing through computer vision and deep learning. The solution is based on Homomorphic filtering, and deep convolutional generative adversarial networks (DCGAN) and works well for less amount data as well. **(Homomorphic filtering, DCGAN, MvTecAD, GAN, synthetic data generation).** **Ticket classification:** In this project, our aim was to classify email based on the semantic information. And sent this email to the appropriate person. We were getting this email into multiple languages. So here we have written language detection, machine translation (Sequence to sequence RNN model), text cleansing, LDA and clustering. **(LSTM, Sequence to sequence, Spacy, NLTK, Text analysis, Stanford-NLP).** **MEDS (India Automation):** MEDS is a data collection product of Morningstar. Every organization has to submit their Annual reports/quarter report to the Stock exchange (SE) in PDF format. Morningstar takes these reports then extracts Face of the statement (Income statement, Balance sheet, cash flow, and Notes) from these reports then fetch data points from the extracted face of the statements and compare these data with earlier reported data. We have Deployed ML models to automatically extract the face of the statement and the model for automatically extract data points from the table then matches these data point to earlier reported morning star data point using machine learning. **(Abby, SVM, random-forest TPOT, Autokeras, Keras, Flask, classification,etc)** **LTI Corp (Travel Claim prediction):** Delivered classification model on employee travel claims based on the data using machine learning (SparkML and SK-Learn). Applied novelty detection, Isolation forest and Box plot (Interquartile range) model to detect outliers using Machine learning **(SK-Learn, NLTK, classification**). **Machine Learning on Mosaic Decisions:**

Mosaic Decision is a platform that serves business intelligence and analytics as a service. Integrated linear regression, decision tree regression models for Regression using machine learning (spark ML).

Integrated Logistic regression, Decision tree classification, and random forest models for Classification using machine learning (spark ML).

Integrated K-Means for clustering using machine learning (spark ML).

Integrated correlation and dimension reduction for data understating and feature importance using machine learning (spark ML, sk-learn, and custom algorithms).

For data preprocessing, integrated label encoder, one hot encoder, outlier treatment, null value treatment using machine learning (spark ML, sk-learn, and custom algorithms)

For text mining integrated tokenizer (simple, regex and n-gram), word2vec, count vectorizer and tf-idf using machine learning (spark ML, sk-learn, and custom algorithms).

Credit Information Report (CIR):

In this project, we hit the Experian information bureau and fetch customer's financial information, based on this information we calculate credit score by a per-implemented scoring logic. It helps a particular entity (bank or financial institute) to make a relational decision on the customer. **(Java, Python, ML).**

Customer Centricity:

This project was the POC for a client. In this project, we hit the bureau and get the important information like contact info, address info etc. Based on this info we trace the person on social media like twitter, e-commerce and try to predict the next possible activities of the person and sent then the best possible offers or suggestions. **(Python, ML, Spark-ML, Stanford-NLP, Spark-Streaming)**

Version Control management system:

At AutoRABIT Inc it was my first project. Here I have done research on version control management systems and integrated almost all major version control systems like GIT, SVN, TFS and Perforce in the AutoRABIT. Now user can do test connection, resister repository, delete a repository, create a branch, delete branch and import branch, etc on the fly at any cloud by AutoRABIT (Java, git, svn, TFS, perforce, git hosted above aws, jQuery, HTML, CSS)

EZ-Checkin:

In this project, we have developed a module that connects Salesforce org to version control and synchronizes data between Salesforce org and version system. User are able to fetch the data according to specific commit, selective data or from head revision and commit to specific repo according to user need (Java, Version controls).

User-Based-AutoCommit:

	User-based-Auto-Commit automatically triggered at the time of CI jobs. It fetches the data from the sandbox or from one org to another org and maintains the history of users based on the users. (Java, Version controls)
Extra-curricular Activities	Attended HiPC (High-Performance Computing) 2014 organized by IEEE and ACM at Cidade de Goa. Participated as a volunteer in ADCONS '13 organized by TEQIP II and IEEE Bangalore at NITK. Participated in two-day workshop on "Recent Developments in Software Testing at NITK. Participated in the workshop on Medical Image Computing at NITK.
Declaration	I hereby declare that all the information furnished above is true to the best of my knowledge and belief.
Date & Place	Pune (Vishal Babu)