VARUNKUMAR AKULA

akulavarunkumar@gmail.com
+1 (984) 810 3922
https://www.linkedin.com/in/varun-akula-945a99a2/

Summary:
- More than 8.5 Years of experience in Extensive software development experience in Quantitative Data Analysis, Machine Learning in Medical, Finance, Ecommerce and Product based industries.
- Well-noted experience in developing predictive modeling, implementing data mining and statistical machine learning solutions to various business problems.
- Experience extracting and manipulating large data sets. Experience translating unstructured, structured, and semi-structured data into actionable insights for decision-making.
- Experience writing code (SQL, Python, Basic R) to extract, clean, analyze and visualize data.
- Statistical Knowledge of Regression models, Chi-squared test, Forecasting and Predictive modeling.
- Experience in text analytics using NLP, Regex on clinical domain.
- Sound knowledge on transformer architecture, having language model application exposures in Word2Vec, BERT, ColBERT, spaCy, GPT etc., with semi-supervised model building experience.
- Experience in building dashboards using Visualization tools like Kibana, Tableau.
- Great understanding of the agile development process.
- Sound knowledge of building and deploying models/Applications through cloud environments.
- Sound knowledge of building and deploying applications through PyCharm, Jupyter Notebooks.
- Developing PySpark scripts as per client requirements. Implementing pre-defined operators such as reduceByKey and CombineByKey in PySpark.
- Written and verbal communication skills with the ability to translate complex issues for a wide audience.
- Designed and managed API system deployment using fast HTTP server. Strong Knowledge on Amazon AWS Services like RDS, AWS Glue(Studio/Ray), Sagemaker, Cloud watch, EC2, Load Balancer using NGINX, S3, Lambda Functions
- Good knowledge in working with application servers - Nginx, Docker containers and Kubernetes.
- Managed teams of 3+ employees and monitored projects to ensure progress(actively involved in Sprint planning and task assignments).
- Ability to work on own initiative and as a part of a team. Willingness to learn new Technologies, openness to new ideas and ability to learn very quickly.

Technical Skills:
Data Analysis | Data Science | AI | Machine Learning | Deep Learning | NLP | Model Development & Validation | Data Visualization | MLOps | GCP | AWS Sagemaker | AWS Services | Application Developments | Python Programming | PySpark | Data Bricks | DevOps | Apache Airflow | GIT | SQL & NoSQL.

Education:
- IIIT Bangalore India, PD Diploma in AI & ML(2021) – 3.4/4 GPA
- Sree Chaitanya College of Engineering (SCCE), Bachelor of Technology in ECE(2013), India**.**

Certificates:
- Data Science Specialization (Machine Learning (Regression, Classification, Ensemble Learning, Random Forests, Xgboost, LGBM), Deep Learning (Tensor Flows, NLP, Neural Networks), IIIT Bangalore, India.
- Deploy Machine Learning Models on GCP + AWS Lambda (Docker). How to Serialize-Deserialize model with scikit-learn & Deployment on Heroku, AWS Lambda, ECS, Docker and Google Cloud – Udemy.

Patients:
- US17344489: Being one of the inventors for MACHINE LEARNING TECHNIQUES FOR PREDICTIVE STRUCTURAL ANALYSIS(LRC) - https://www.credential.net/421dc8a5-e712-41f6-97d9-ae79232ccbfc#gs.jv8i60

Professional Experience:

United HealthCare   **08/2017 to Present**
Sr Python Developer / Data Scientist

Responsibilities:
- Working on Automating the Appeal & Grievance of Preprocess for Medical Claims. Objective is to predict accurately for better processing of medical claims in a short duration.
- Involved in machine learning and optimization workflow which contains preprocessing, modeling and postprocessing.
- Train CNN for over 10 Million vectorized documents to identify different variations of pages across the business, and to train a separate classification model to identify few of the business fields from the vectorized documents.
- Also, build a Template Vectorizer on top of SkLearn's Count-Vectorizer, that could identify various templates across the document corpus
- Build Reinforcement-learning algorithm to bump up accuracy.
- Create custom NLP algorithms and annotators to evaluate medical record data.
- Using Text Vector to vectorize the contents of the new documents and then use pre-trained CNN and classification model to identify the type of page in a doc and business fields respectively. Connection with SQL DB and Oracle DB for business fields' validations.
- Working in Natural Language Processing (NLP) for Text Classification & Topic Modelling and Computer Vision for unstructured data processing.
- Created a class for different algorithms in Python in a virtual environment for forecasts to consume. Expertise writing codes in python from scratch with minimal intervention of existing libraries.
- Developed a predictive model to predict the right label using machine learning algorithm and performed various optimization techniques to standardize the file.
- Building & Deploying models on AWS cloud using SageMaker. Maintaining & Managing different environments (DEV, STAGE, PROD) in Cloud(AWS/AZURE EC2) as well as Standalone linux machines extending them with Docker and kubernetes tools.
- Creating & Exposing Dashboards to show insights and reports to users using both Kibana and Tableau. Configuring Interactive Dashboards with Drill-down capabilities, filters.
- Created various visualizations using Tableau for Claims/Grievances Analytics.
- Participated in features engineering such as feature intersection generating, feature normalization and label encoding with Scikit-learn preprocessing.
- Used Python 3.X (numpy, scipy, pandas, scikit-learn, seaborn) and Spark 2.0 (PySpark, MLlib) to develop a variety of models and algorithms for analytic purposes.
- Worked on stored procedures to create a custom ensemble algorithm which uses weighted average concept to calculate metrics from different kinds of user-defined combinations from the existing models in the system.
- Built regression models like Lasso, Ridge, SVR and Xgboost to predict Customer Lifetime Value. Built classification models like Logistic Regression, SVM, Decision Tree, and Random Forest to predict Customer Churn Rate.
- Used F-Score, AUC/ROC, Confusion Matrix, MAE, and RMSE to evaluate different model performance.
- Used interactive visualization using Bokeh, Plotly and PowerBI for Visualization purposes at different levels for different purposes.
- Communicating with other teams and business in effective delivery of project requirements. Guiding and managing team members .

Environment: Python/R, Python 3.x (Scikit-Learn/SciPy/NumPy/Pandas/Bokeh/Matplotlib/Seaborn), Spark (Pyspark), SQL, Oracle, MongoDB, Microsoft SQL Server Management Studio, Machine Learning (SVR, RandomForest, LGBM, XGBoost), DeepLearning(CNN, RNN, Transformers), Azure , AWS, AirFlow, Kibana, Flask framework, Agile/SCRUM.

HashInclude Computech Pvt Ltd, Bangalore, India   **10/2016 – 07/2017**
Software Engineer

Responsibilities:
- Experience in using collections in Python for manipulating and looping through different user defined objects.
- Used Unit Test Python library for testing many programs on python and other codes.
- Collaborated with the team of QA to build and populate DB and to ensure the standards.
- Implemented RESTful Web-Services for sending and receiving the data between multiple systems.

- Created Business Logic using Python to create Planning and Tracking functions.
- Heavily leveraged Python & graphics APIs for creating graphics and serialization libraries for encoding data in XML/JSON formats.
- Writing REST APIs, as part of developing web-based applications for insurance premium calculations, using Django and REST framework
- Responsible for using the AWS tools like RDS, Cloud watch, EC2, Load Balancer using NGINX, S3, Lambda Functions.
- Managed website tagging and tracking to inform ecommerce performance and strategies.

CoBot Systems(Sister Company of HashInclude), Bangalore, India   03/2014 to 10/2016
Software Engineer

Responsibilities:
- Single handedly created UI and Backend for the projects. Actively involved in the complete development cycle that is from the Design Document, defining hierarchical DB structure, backend services.
- Collaborated with data engineers and the operation team to implement the ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
- Time Series Analysis for Forecasting on Sales, RO, Services, Leads Opportunities, Performed Document classification on vehicle sales using ML algorithm. Performed EDA, different statistical techniques on data to drive insights and solutions on it.
- Designed rich data visualizations to model data into human-readable form with Bokeh, Seaborn and Matplotlib.
- Provided expert technical guidance to the team on modeling, data analysis, and programming.
- Developed Restful services for returning analytic results, Implemented web applications in Flask frameworks following MVC architecture. Created APIs, database Model and Views Utilization Python to build responsive web page applications.
- Implemented user interface guidelines and standards throughout the development and maintenance of the website using the HTML, CSS, JavaScript and jQuery.
- Design and implementation of low-latency, high-availability, and performant applications Implementation of security and data protection.
- Extensive use of Beautiful Soup and Python Regex for performing Web-Scraping tasks.
- Explored different implementations in the python environment for data ingress and summarization by using packages like Pandas, Numpy and Scipy.
- Creating custom ETL Process for normalizing data from various sources like CSV, Excel, JSON, XML etc.
- Optimizing and transforming SQL queries and using No-SQL DB like Mongo-DB, ElasticSearch.
- Deploying applications and migrating existing on-prem implementations onto cloud based solutions like EC2, RDS, SQS, Lambda Services, ElasticSearch, S3 Glacier and Amazon API Gateway.

Environment: Google Cloud Platform, AWS, Big Query ML, Kube Flow pipelines, Kubernetes, Flask Framework, Python 3.x (Scikit-Learn/SciPy/NumPy/Pandas/Bokeh/Matplotlib/Seaborn), SQL, Oracle, MongoDB, Machine Learning (XGBoost, SVR), .