MYNAMPATI PRUDVI
E-Mail: Prudvireddy20@gmail.com
Contact No.: 9573412468
Github Link :https://github.com/mynampati-prudvi
LinkedIn:https://www.linkedin.com/in/prudvireddy20/

Career Objective:

Seeking a challenging position, with a strong emphasis on Machine Learning, Data science and Hadoop where I can use my abilities and explore the best of skills and acumen to become a valuable asset to the organization.

Profile Summary:

- Over 3 years of experience in the IT industry including Machine learning, Big Data-Hadoop, Data Warehousing technologies and Data Science.
- Image recognition for notifying the person's name who unlocked or locked the smart home doors.
- We take the data and process the data insides to find the delay reasons and Accurate predictions for the Arrival and Departure flights.
- Arm and Disarm the Home security system using face recognition.
- Using CNN (convolution neural networks) Train the Images to recognize a person.
- Using the Facenet package, find whether the person who appears is real or fake.
- Data visualization for Business intelligence problems using Tableau.
- Data preparation, data exploration and data manipulation of huge datasets for data analysis.
- Openrefine tool is used to filter the unwanted and missing values or outliers etc.. data preparation process.
- Regression and classification algorithms for future prediction.
- Google Assistant Smart home (Z-wave, SRF and Power G Sensor) device action control using voice commands.
- Cassandra api calls used in rest to retrieve data for Google Assistant and Amazon Alexa data source.
- AAA security layer is developed for user validation credentials using keycloak.
- Maria DB is Used to score credentials as backend for Keycloak and Nodejs for Rest calls.
- Real time experience in mysql query language to retrieve or check the status of the data.
- Apache kafka is used as a message broker which validates user credentials using keycloak server.
- Hands-on experience in configuring, developing, analysis, monitoring, debugging and performance tuning of Big Data applications using Hadoop
- Collecting amounts of log data using Apache Flume and aggregating using PIG/HIVE in HDFS for further analysis
- Worked with customers closely to provide solutions to various problems
- Quick learner and self starter to go ahead with any new technology.

Education Details:

Qualification	Institution	Board/University	Year of Passing	Percentage
PGP of Data science and Machine Learning	University of Chicago Graham School (Jigsaw Academy)	University of Chicago Graham School (USA) Online	2019	95
B.Tech [Electronics & Communication Engineering]	Sri Indu College of Engineering &Technology	JNTU, Hyderabad	2016	71.90

Technical Skills:

- Big Data/Hadoop ecosystem : HDFS, MapReduce, HIVE, Pig, Sqoop,Flume ,Athena,Glue,Lambda.
- DB Languages : SQL, MYSQL.
- NoSql : Cassandra
- Operating Systems : VMware, Linux, WINDOWS.
- Languages : Python, Nodejs.,HTML,CSS.
- Tools Data Preparation : Open refine, Excel,Cognito,Api Gateway.
- Algorithms : Linear Regression, Logistic Regression, Decision Trees, Random Forest, Clustering ,CNN,SVC, Extra Trees.
- Cloud Environment : Google Cloud Platform, AWS
- Deployments : Kubernetes , Docker .
- Storages : Aws Data Lake.

Certification:

- **Data Science and Machine Learning** From **University of Chicago Graham School (Jigsaw Academy)**
- **Diploma in Hadoop Big Data from NIIT, Hyderabad in 2016**

Professional Summary:

Project-1:
Organisation: Zestiot Private Limited
Project: Storing the data into DataLake

We capture data coming from the IOT devices to the Data Lake for analytics.

Responsibilities:
- Here we capture all the data coming from the IOT devices Fitted to the equipment in the Airport which serves the flights froms its Arrival to Departure.
- The data from devices is formed into one structure format and it is stored in Mysql DB using a processing layer.
- We use **AWS Glue** to retrieve the data from **Mysql** and store in S3 DATA LAKE.
- For analytics we retrieve the from **DATA LAKE** using **AWS Athena** to **POWERBI** for analytics

Project-2:

Organisation: QOLSYS SOFTWARES INDIA PRIVATE LIMITED:
Project:AAA SERVER

We use AAA(AUTHENTICATION,AUTHORIZATION & ACCOUNTING) server as security for client and cloud for that KEYCLOAK for protecting data in cloud.

Responsibilities:

- **KEYCLOAK** is a tool we use for protecting data on the cloud for accessing registered users only.
- KEYCLOAK is authenticated by generating a token by the webserver on top of undertow.
- It can set according to the roles and policies allocated for the client or user.

- It is set up in round robin cluster format for the request to share the load.
- **MARIA DB** is the database used for the storage of keycloak data about users, clients.
- **HA-PROXY** we use for load balancing requests from the 3rd party client.
- Jmeter is used as a testing tool of the current working tool for a number of Authentication and Authorization requests.
- **REST SEVER** on top of undertow to make requests processed to the keycloak for the authentication.
- **KAFKA** acts as a message broker according to the topics requests from the 3rd party users (producers and consumers).
- **CASSANDRA** is a database used to store the total data generated during the entire process and it can be used for future analytics or recovery.
- **Dockerize** the **Keycloak Server** and deploy it in **Kubernetes**

Project-3:

Organization : QOLSYS SOFTWARES INDIA PRIVATE LIMITED:
Project : Google Assistant :

To make life easier with advanced technology using google Assistant voice commands to function home appliance.

Responsibilities:

- My company is a product based Home Security System in the USA.
- At the initial stage of development of google assistant we need to have an idea about the functionality of Z-Wave , SRF , Power-G device with the Qolsys kit.
- **Node Js** is a programming language which helps us to interact with your device and Google Assistant.
- Developed an Authentication barrier Keycloak for the user to access Qolsys kit (panel) from the Google assistant home kit app .
- This process goes in 3 steps (**Sync, Query, Execute**) from the user mobile app (Google Assistant) to the device like Light, Thermostat, mart Socket, Dimmer, Fire Sensor etc.
- **Rest layer** which is authenticated with **Keycloak server** which helps us to query the **Cassandra DB**
- When request for Sync we pick the device data from the **Cassandra DB** which are connected to the respective User and Panel (Qolsys it)
- Report State plays a key role for this Google Assistant like whether the current state of the device is "ON " or "OFF", Report State stores the status of the device previous action and response whenever we request for the status.
- All the Report State status stores in the Home Graph DB of google.
- Generate a **War** file and **Dockerize** the application with the app properties and deploy as a **pod** in **Kubernetes**

Project-4:

Organisation : QOLSYS SOFTWARES INDIA PRIVATE LIMITED:
Project : Image Recognition:

Notifying a person who unlocked or locked the door,who all are trying to open the doors,etc with respective names to the master user registered in the Qolsys kit will receive a message alert.

Responsibilities:

- Machine learning Tensor flow is the good fit for solving this kind of problems like face detection, Image classification.
- Tensor flow is google api for the image classification available.
- Qolsys Home Kit has a front camera and it is used for video streaming.
- Using that video streaming data to capture the face images who stands in front for locking or unlocking any device, Our model which deployed in the panel will read that data and send a message to the master user.
- We take a maximum 5 people video for 30 sec of their face in all the angles and we capture in picture format and make a folder for that image.

- All images which are stored are separated into train, test and validate sets.
- Train dataset is trained using **KNN** and **CNN (Convolution Neural Network)** with TensorFlow and keras to generate a model.
- Verify the Accuracy of the model Predictions using the test data images and their accuracy.
- Model is deployed into the panel and using the front camera installed in the panel will predict and notify using a message.

Project-5:

Organisation : QOLSYS SOFTWARES INDIA PRIVATE LIMITED:
Project : Data Analytics:

Here I collect the data file from cassandra and filter the data according to the requirements.

Responsibility :
- I read the data into a jupyter **notebook** which is in csv file format.
- Here we need to check the connected and disconnected panel with the cloud server.
- we need to notify the users that they lost their connectivity with the network through mail and mobile.
- We store the event logs to the **HDFS** in **Google Cloud Dataproc**
- We use **Hive** to query the hive database where the data is loading from HDFS to cross verify the problem and we need to suggest the resolution for the cause .

Data & model working Experience

Since 2018 sep '28 with University of Chicago Graham School(Jigsaw Academy)

Key Result Areas:

Project Title : **Cost Optimization on products.**

Problem Statement

- What should be the price at which the products (class / group of products) should be sold? (Need to support "finding an optimal price point for every SKU will drive higher top line and bottom line benefit."
- What factors define and drive prices of products?

Goal:
- Optimal price of sale given a product/group of products along with its confidence interval.
- The impact of variation in variable on pricing should also be explained.

Responsibility:
- We need to gather the data to point of storage and we need to understand the structure and insides of the data.
- We use **Jupyter Notebook** to read the files.
- Checking the Null values ,Removing duplicates ,Rename column for better understanding ,Remove outliers
- We need to explore the insides and important points using the **Tableau**.
- we need to select one target variable from the given data and rest are considered as independent variables
- Here the target variable is continuous. We choose supervised Learning with a regression problem.
- We pass the data to different algorithms like **Linear Regression,KNN,Decision Trees,Random Forest,Lasso and Ridge regression**.
- To know the depth of the **Decision Tree or Random Forest** we choose **Grid Search Hyperparameter** to choose the depth.
- By considering the **Train and Test Accuracy** we choose the model is **over fitting or under fitting** to the model and tune the parameters .
- For **Linear Regression** we follow **forward or backward policy** to filter the data which are co-related and not adding weightage to the train data.
- To check the Accuracy score we choose **R2 score,Accuracy,ROC curve,Mean Square Error,Absolute Square Error**.

Project Title: News Articles Classification using Machine Learning

Problem:

- The Global News Group has 1000s of news items being fed daily to its news aggregation centers
- They have a challenge to classify the news articles into various categories such as "Sports", "Entertainment", "Politics","Business" etc.
- It is currently a manual time consuming process that is impacting their business .
- Solution : Supervised Machine Learning Algorithms and models, of course.

Goal:

- Predict creates a text classifier that will streamline the process of categorizing news publications.

Responsibility:

- First we need to understand the Business logic and Requirements.
- We need to gather the data to the point of storage and we need to understand the structure and insides of the data.
- We use **Jupyter Notebook** to read the files.
- Checking the Null values ,Removing duplicates ,Rename column for better understanding ,Remove outliers
- We need to explore the insides and important points using the **Tableau**.
- we need to select one target variable from the given data and rest are considered as independent variables
- Here the target variable is categorical; we choose **supervised Learning** with **classification** problems.
- We pass the data to different algorithms like **Logistic Regression,Decision Trees,Random Forest,Voting Classifier,Bagging**.
- To know the depth of the **Decision Tree or Random Forest** we choose **Grid Search Hyperparameter** to choose the depth.
- By considering the **Train and Test Accuracy** we choose the model is **over fitting or under fitting** to the model and tune the parameters .
- To check the Accuracy score we choose **Log Loss,Accuracy,ROC curve,Mean Square Error,Absolute Square Error**.

Declaration:

I hereby declare that the above particulars are true and correct to the best of my knowledge.