

K PAVAN KUMAR

Email: pavvankanamarlapudi@gmail.com

Contact: +91-8019620741

OBJECTIVE

An Urge to work in an organization to efficiently leverage my skills gained through my experience and learning for better decision-making support for driving business growth

PROFILE SUMMARY

- About **6+ Years** of experience in Python Programming/Data Science.
- **Google Cloud Certified** Professional Data Engineer.
- Currently working with **HSBC Software India Pvt. Ltd** as **SSE** in Hyderabad.
- Pursued **DATA SCIENCE SPECIALIZATION** from **JIGSAWACADEMY**.
- Pursuing **BIG DATA SPECIALIZATION** certificate program from **JIGSAWACADEMY**.
- Experience in **Python Programming, Machine Learning, Statistics, Regression-Linear, Logistic**
- **Tool and Techniques worked on:** Supervised and Unsupervised classification (Naïve Bayes, Support Vector Machines, Random Forest etc.), Regression.
- **Interest Areas**-Mathematics, Programming, Automation, Business Analytics, Data science, Statistical Modelling, Predictive Modelling, Text Mining, Machine Learning.

QUANTITATIVE SKILLS

<ul style="list-style-type: none">• Logical and Analytical abilities	<ul style="list-style-type: none">• Problem Solving skills
<ul style="list-style-type: none">• Leadership	<ul style="list-style-type: none">• Willingness to learn new things and apply
<ul style="list-style-type: none">• Domain Knowledge: Manufacturing/Retail.	<ul style="list-style-type: none">• Self-Starter and taking initiatives in building my own skills

ANALYTICAL SKILLS

<ul style="list-style-type: none">• Statistics and Predictive Modelling- Linear and Logistic Regression, Hypothesis Testing, ANOVA.	<ul style="list-style-type: none">• Machine Learning- Supervised and Unsupervised Learning• NLP
<ul style="list-style-type: none">• Data Visualization- Matplotlib, Seaborn, ggplot2(R)	<ul style="list-style-type: none">• Deep Learning (Keras)

Pavvankanamarlapudi@gmail.com

+91-8019620741

TOOLS

<ul style="list-style-type: none">• Python- Anaconda, Pandas, Scikit- Learn, NumPy, pyspark, selenium, Keras• Version Control System (Git & GitHub)• HTML, MS Office.• Robotics Process Automation (UiPath)	<ul style="list-style-type: none">• Packages used- dplyr, ggplot, tm etc.• R Project, R studio- R for Data mining and Analysis• Tableau• Postman (API Testing)• Flask & Django Rest Framework
--	---

EDUCATION

Degree/Course	Institute/University	Year	Percentage
Data Science Specialization	Jigsaw Academy	2016-2017	-
B. Tech (Electrical & Electronics)	JNTUK, Kakinada	2012-2016	82.07%
12 th	Sri Gayathri Jr College	2010-2012	93.40%
10 th	G.C.S. S Jr College	2009-2010	93.00%

CERTIFICATIONS

- Certified as **Google Cloud Professional Data Engineer** from **Google**.
- Certified as **Advanced RPA Developer** from **UiPath**.
- Certificate course from **JIGSAW ACADEMY** on **DATA SCIENCE..**
- Completed Certificate course from UDEMY on **PYTHON BOOTCAMP** and **Pyspark**
- Completed Certificate course from UDEMY on **Complete MySQL for Data Science**.

PROFESSIONAL EXPERIENCE

HSBC Software India Pvt. Ltd. Senior Software Engineer Dec-2021 to till date.

Project: **Data Migration to Cloud Program (On-Prem to GCP Data Warehouse)**

Role: Data Engineer

- Ingesting the data from various sources to RDL (BQ) using proprietary tool Juniper.
- Interacting with ITSO and SME's to gather the source details.
- Automated the entire process for creation of prerequisites for Ingestion process such as Ingestion sheet Preparation, DDL Queries preparation and BQ Load Scripts Preparation
- Automated the Config files creation which help to provide as Input to Job running in Juniper.

Pavvankanamarlapudi@gmail.com

+91-8019620741

- Worked on *Batch Pattern* and *Real time Pattern* of Moving data to cloud.

Tools/Techniques: Python, Pandas, VBA Macros, GCP Big Query, Cloud Data Flow, Pub/Sub, Kafka, GCS, VSCode

Project Match Engine Builder (In-House application)

Role: Python Developer (Automation)

- Application needs to evaluate the risk profile of the customers of HSBC if any requirement needs to transfer the Data across the borders
- As part of Risk evaluation, Project initiators has to raise a case in the application so that corresponding jurisdictions, countries involved people will review and approve it
- Build a matching logic if user case matches with already raised case so that the TAT for approval of the case will come down by auto approving the same.
- Data Visa is a HSBC Managed in-house app to track the control the legal & Compliance issues through the Cases while Data being shared across borders
- To build the match engine for checking the new case which if exists in the system (Data Visa)
- If exist, to derive the match percentage with recent historic Case and produce report with all attributes that are matching and un-matching.
- The API's has built for the POC using Python-Flask Framework and Integrated with UAT Data Visa application
-

Tools/Techniques: Python, Pandas, SQL Server, Postman (API Testing), Flask, HTML, CSS (for Reporting), SMTP, VSCode

Project: Data Privacy

Role: SQL Developer

- As part of application maintenance, user will raise the request to change the backend values.
- As per the user request analysis has to be done and prepare corresponding sql queries to modify the data in backend production tables so that it reflects on Front-End application and they will proceed further.
- Expanding it to advance level to auto generate the queries based on the user request and past historical queries using Lang Chain Framework and GPT Models.
- TO predict the query as per the user problem and customize it based on the problem statement more.

Tools/Techniques: SQL Server, Stored Procedures, Triggers, Python

Project: **Strategically Determine Surged Price Pattern via Predictive Data Analytics**

Client: **Lam Research Corporation, USA**

Role: **Data Scientist**

- Optimized manufacturing cost reduce complexity and enable strategic decision making with the data-driven predictive analytics.
- By solving business problem through the combined power of Data Science and Machine Learning techniques. We analyzed the behavior of each supplier, as the price surge / extra fees was incrementing year-over-year (YOY).
- Moreover, we trained the predictive model (developed in Azure ML Studio) on historical data of purchases for last three years, which consists of purchase order data, supplier features and historical transactions, product features etc.
- This real-time estimation helped the client in reducing the expedited delivery expenses.

Tools/Techniques: R Language, Python, Pyspark, Jupyter Notebook (anaconda), Machine Learning Models (Regression), Random Forest, Azure ML Studio (Deployment)

Project: **Predicting Return Behaviour of Customer after Repurchase using ML Techniques**

Client: **Macy's INC, USA.**

Role: **Data Scientist**

- It has been observed that some customers are re-purchasing a merchandize (which they have purchased sometimes back with a higher price) with an "Intent to return the repurchased merchandize" to get a "price adjustment" against the "merchandize purchased earlier".
- The goal is to develop a solution that will offer a "proactive price adjustment" suggestion if the new transaction is a "Repurchase with an intent to return for price adjustment" pattern.
- For learning developed entire project workflow in Pyspark.

Tools/Techniques: R Language, Python, Pyspark, Jupyter Notebook (Anaconda), Machine Learning Models (Classification)

Project: **Chabot (B2C)**

Client: **Bluebird Inc**

- An automated response machine to instruct users to solve basics problems/incidents related to **Bluebird devices** by understanding contexts.
- **Benefits:**
 - o This reduces (but not eliminates) dependence for incident resolutions, as well as can reduce number of services now tickets

Features:

- o Context Based responses
- o Text and Image Instructions for better understanding
- Sample user queries implemented:
 - o Blue Bird has Physical Damage.
 - o Blue Bird/How to Reboot
 - o Blue Bird - Low Battery Symbol

Tools/Techniques: Slack API, Python, Flask, ngrok, Naive Bayes Algorithm

Project: **Domain based NLP Chatbot for DevOps Operations**

Client: **Kohl's INC Ltd.**

- Worked as Developer to build a Rule-based & NLP based Multi-User Chat-bot for DevOps Team to reduce the internal manual work.
- The client's technical support team needed to input commands multiple times for fetching similar information from the system.
- As a result, there was high mean time to resolution (MTTR). The client wanted to automate this process to improve productivity, better resource utilization, and quicker root-cause analysis.
- Reduction in time required for resolution and RCA (root cause analysis)
- The number of tickets raised by the client's support team for fetching information reduced by **40%**.
- Getting Dockers Logs, check health of applications, Restarting Dockers, IAM Authentication, creating incidents are some of the use cases implemented.

Tools/Techniques: Cisco WebEx API, Python, Flask, Ngrok, Naive Bayes Algorithm, Tonomi API, Shell Scripting, JIRA API

Project: **Autobot**

- Developed NLP Based Autobot.
- Mimic ticket handling in production Environment.
- Based on incident raised by Support/Production Team, Classifying the Incident and fetches the

Pavvankanamarlapudi@gmail.com

+91-8019620741

required information from JIRA API.

- Assign the Detected Incident to corresponding person as per predefinedRoles.
- Auto closing the issue as soon as assignee approves it.

Tools/Techniques: Cisco WebEx API, Python, Flask, ngrok, Naive Bayes Algorithm, Tonomi API, Shell Scripting, JIRA API, Cron Schedulers

Project: **CR Inventory Management using Predictive Analytics**

Client: Charrles Routte

Role: Data Analyst

- Performed EDA using HIVE Queries to gain valuable insights from available sales data.
- Forecasted Daily and weekly Sales to maintain the inventory stocks as per predictions.
- Finally, mapped the predicted inventory with weekly Forecasted sales.
- Implemented Workflow in Azure ML Studio.

Tools/Techniques: Python, Jupyter Notebook (Anaconda), HIVE, Tableau, Azure HD Insight

Project: **Web Scrapping Using Python**

Client: Kohl's INC Ltd

Role: Python Developer

Scrapping Live Data from Azkaban Job Scheduler

- Scrapped the live data generated through Big Data Map Reduce work via Azkaban Scheduler UI using python Selenium.
- Generating Summary reports of all jobs running day from scrapped data.
- Sending mails automatically based on predefined frequency to clients using Windows Task Scheduler.

Tools/Techniques: Python, BeautifulSoup, Selenium, Jupyter Notebook, Windows Task Scheduler, SMTP

Monitoring the Recommendations of Shopping Website

- To monitor the performance to **Recommender system** for every **15** mins built on Technique.
- Scheduled script to get the displayed recommendation on a webpage from specific channel (Mobile, Tablet, Website) on hourly bases.
- Validating if recommendations are properly being served on corresponding channels.
- Made Automatic mail delivery system through python to send the scrapped recommendations displayed on webpage to the client.

Tools/Techniques: Python, BeautifulSoup, Selenium, Jupyter Notebook, Windows Task Scheduler, SMTP

Project: **BEAT (Big Data and ETL Automation Tool)**:

Client: Nike Inc Ltd.

Role: Sr. Big Data Engineer.

- Has been working as Backend Python **Django/Pyspark** Developer
- Automating ETL Workflow validations by generating executable queries
- Fetching the Data from various RDBMS Systems or similar databases such as **Big Query (GCP), MySQL, SQL-SERVER, Amazon S3** using python connectors
- Profiling the data in various aspects as result to perform the Quality Check of the Data before and after ETL migration.
- Report generation by consolidating all QA outcomes and triggering the alerts to Slack or Outlook.
- JIRA Integration to raise the bugs, defects from BEAT to track the issues.
- Transformed UI Design to Utilities as a python package which deployed across all prod clusters to validate larger volumes.
- Sending failure alerts to Slack Channel and Mails using SMTP and API's
- Created DAG's using Airflow to schedule the quality check in ETL Pipelines.

Tools/Techniques: Django Rest Framework, Pandas, Pyspark, Apache Airflow, Jenkins, Git/Bitbucket, MySQL, Amazon S3, EMR, Cerberus Client, JIRA API, Confluence API, Slack API, SMTP

EXTRA CURRICULAR ACHIEVEMENTS

- 10th class school **Topper**.
- Received Gold Medal for achieving **100%** in Science in S.S.C
- Recognized as the best performer and received a **Shining Star Award** for the year **2019-2020** for building a **NLP based Domain Chabot**.
- Recognized as **Star Performer** for the year 2021-2022 (Q1) for automating the whole process of Ingestion as part of Data on Cloud Program

PERSONAL DETAILS

Father's Name	:	K. Srinivasa Rao
Hobbies	:	Coding, Painting, listening to music, and Reading books.
Languages Known	:	English, Telugu.
LinkedIn profile	:	www.linkedin.com/in/pavan-kumar-a46b7897
Git-Hub	:	https://github.com/pavvankanamarlapudi/pavvankanamarlapudi
Website	:	https://analyticswithr.weebly.com
Medium	:	https://medium.com/@kpavankumar_19821
Address	:	D. No: 12/25/101, Beside Kotak Mahindra Bank, Above Venkataramana Eye Care, Kothapet, Main Road, Guntur (A.P)-522001.

Pavvankanamarlapudi@gmail.com

+91-8019620741

DECLARATION

I do hereby the declared above information is true to the best of my knowledge and belief.

Place: Guntur,

Date: -----