|  |  |  |  |
| --- | --- | --- | --- |
| **AMBER KATARIA** | | | |
| **CONTACT**  **Address:**  WZ- 273A, Street no.-3, Gurunanak Nagar, Tilak Nagar, Delhi – 110018  **Phone:**  +91-7678141380  +91-9711399322  **Email:**  [katariaamber@gmail.com](mailto:katariaamber@gmail.com)  **EDUCATION**  **2012- 2016**  **•** B.tech - Electrical and Electronics Engineering, Maharaja Agrasen Institute of Technology, GGSIPU, Delhi  **2011 - 2012**  **•** Senior Secondary, K.V. ,Sec-8, Rohini, Delhi  **2009 – 2010**  **•** Secondary, K.V. ,Sec-8, Rohini, Delhi  **SKILLS**   * Machine Learning * Python * Data Visualization * SQL * Matplotlib * Pandas * NumPy * Flask * SciPy * Data Mining * AWS * GCP * Power BI * Google Data Studio   • **Tools**  - Jupyter Notebook  - Pycharm  - Power BI  - Google Data Studio  - Tableau  - VS code  - SQL  **TRAININGS AND CERTIFICATIONS**   * Coursera - Applied Machine Learning in Python * Coursera – Applied Plotting, Charting & Data Representation in Python. * Coursera – Introduction to Data Science * Coursera – SQL for Data Science * Coursera – Programming for Everybody-Python * **TruLabz, Rohini:** EMBEDDED SYSTEM MICROCONTROLLERS: 90 days training   program. It includes AVR, ARDUINO, PIC and  8051.  **ADDITIONAL SKILLS AND TOOLS**  • Programming Languages  - Python  - Embedded C  - C++  • Embedded Systems  - AVR  - UART  - SPI  - Arduino  - Raspberry Pi  • Tools  - Proteus ISIS Professional  - Atmel Studio  - AVR Khazama  - Autodesk 123D design  - Service Now  - SharePoint  **HOBBIES**  • Swimming  • Origami  • Travelling  • Teaching |  | **OBJECTIVE**  Data Scientist familiar with gathering, cleaning and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, algebraic and other analytical techniques. Highly organized, motivated and diligent with significant background in Data Science.    **PROFESSIONAL SUMMARY**   * Having almost 3.2+ relevant years (Total 4.2+) of experience as a Data Scientist. * As a Data Scientist, skilled in Machine Learning, Python, data analytics, SQL, flask, Pandas, numpy, PowerBI, Google Data Studio, AWS, Linux, I identified and integrated new datasets that can be leveraged through product capabilities. * Automated the processes using Python and Machine Learning Models like Decision Tree, Random Forests (Classification and regression), K-Means Clustering, KNN, Linear and Logistic Regression Methods. * Data Wrangling, EDA, selecting features, building and optimizing classifiers using machine learning & Statistical techniques. Identify relevant data sources to mine for client business needs and collect large structured/unstructured datasets and variables. * Good understanding of Statistical Modeling, Multivariate Analysis and Standard Procedures. * Worked in stats function with Numpy, visualization using Matplotlib and Pandas for organizing data. * Devise and utilise Machine Learning algorithms and models to mine data stores, perform data and error analysis to improve models, and clean and validate data for uniformity and accuracy. * Analyse data for trends and patterns and interpret data with a clear objective in mind using PowerBI and Google Data Studio. * Implement analytical models into production by collaborating with other data scientist team members. * Client collaboration & requirement gathering by working with economists, business stakeholders and different teams at ground to successfully develop & deliver simplified analytical Renewable Resource Management models * Solutions are deployed at AWS and now migrating it to GCP.   **WORK EXPERIENCE**  **Applied Solar Technologies India Pvt Ltd, Delhi 09/2020 - Present**    ***Data Scientist***  ***Projects***  **SRMD – Predictive Fault Analysis**   * Project description: SRMD - Faults are spontaneous and hard to catch as these faults can occur in any sensor of BTS system of SRMD types of Telecom Towers which leads to site down and penalties of millions on our organisation for not maintaining the uptime of 99.99% as per SLA. The goal of this project is to "predict the faults" based on the behaviour change of raw string of sensor data. * To reach this goal, we are getting data from ONM box and SRMD box, in the form of unstructured data which is transformed to Structured data in MSSQL server. We used our SRMD box sensor data from Battery Sensor, Diesel Generator Sensor, Battery Voltage Sensor, Electricity Sensor, received date & time and Latitude & Longitude of Site. * Loaded data into Jupyter Notebook from MSSQL for data cleaning and wrangling. * Utilized various techniques like Histogram, Bar plot, Pie-Chart, Scatter plot, Box plots to determine the condition of the data. * Conducted EDA(Pandas) to look for trends, patterns, grouping, and deviations in the data to understand the data diagnostics and process the data by merging, finding outliers, errors, trends, missing values and distributions in the data. * Developed range of predictive models to identify the faults in the equipment using Logistic Regression / Random Forest /Decision Tree in Python. * Generated an average accuracy of around 83% across zones using scikit (Machine learning) library of Python. * Worked on validation model and analyze ROC,AUC score,Accuracy, precision and recall value. * Environment: Jupyter Notebook,PyCharm, SQL-Server, Microsoft Excel   **Energy Saving Telecom Sites Analysis – Sarai Chhabila**   * Project description: This problem statement is targeted to understand the changes and trends, majorly, in Solar Energy synthesis over the year w.r.t. to weather change, money spent on diesel and Reduction in diesel theft after the R and D work done on the sites. * End to end project management and solution development from requirement gathering, data preparation, data wrangling, data churning, data blending and reporting the benefits of the pilot within the business. * Creating PowerBI dashboard to show to the higher management as per the identified Key Performance Indicators (KPI's) among all the given attributes   **Google Data Studio Dashboard –- Solar Health Management**   * Project description: To create and share engaging reports and data visualizations of Solar health status of the sites for senior management and clients. * Transform raw data into the metrics and dimensions needed to create graspable dashboards to show Solar Energy Synthesis in KWH, Solar Panel Life vs Solar Energy Synthesized graph, Battery health status and boost charge requirement. * Deployed automated dashboards that update regularly to our web portal and also enabled weekly auto email to arm Senior Management, Clients and other teams with the knowledge of key metrics, to compare, filter and organize the exact data needed on the fly in one report   **EnGeniusLab, Delhi 11/2017 - 08/2020**    ***Data Scientist***  ***Projects***  **HealthCare Provider Fraud Detection Analysis**   * Project description: Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims.The goal of this project is to “predict the potentially fraudulent providers” based on the claims. * Converted raw data to processed data by merging, finding outliers, errors, trends, missing values and distributions in the data. * Accomplished Data analysis, statistical analysis, generated reports, listings and graphs Instigated the Test Analysis to understand the potentiality. * In this Project, we have used machine learning algorithms (LR/DT/RF) to classify fraudulent behavior of Healthcare providers. * Our models consistently performed with ~0.88 Accuracy, ~0.80 AUROC score and ~0.55 Kappa Score. * Environment: Python,Jupyter Notebook,Spyder, SQL-Server, Microsoft Excel   **Alexa Skill – Virtual CEO**   * Alexa skill hosted on Python to present a virtual CEO for all business meet ups and other gatherings, respond accordingly with the voice of Mr. Mahashay Dharampal Gulati (CEO, Mahashian Di Hatti Private Limited). * Key Skills used: Python, AWS, Lambda, Audio Player Interface Reference, NLP   **Christopher- A smart home assistance (Like Alexa)**   * Christopher is a cloud-based voice service platform and virtual assistant that powers an entire smart device ecosystem coded in python.   **Software Used:**  Pycroft OS , Putty, Xming X11, AlsaMixer  **Hardware Used:**  Quadcore ARM Cortex-A53, 64Bit - Raspberry Pi3b+,USB Speaker with impedance 4 ohms**,** PS3 eye – Mic, CAT 6 LAN cable with RJ45 (ethernet) socket  ***Sr. Embedded System Developer***   * 1 year of experience in developing various projects for clients by analyzing, proposing and implementing solutions for Business Problems while working closely with the technical as well as management teams to deliver the product in economical, fast and effective way. * Build robust embedded software, drivers and frameworks to run on IOT devices. Experienced in boot loader, device driver, projects including real time connectivity, device fault management.   ***Projects***  **Talkative Robo-Waiter**   * Atmega328p, Arduino IDE, USART, 3D printer, Mic and speaker sound card with LM386 IC, Metal Gear 150RPM motors, LED, Speaker, Embedded C, SD card Module   **IoT based Smart Health Care Monitoring System**   * This project includes hardware which will measure blood pressure, heart rate, weight & temperature, and send to Android application & MySQL Database. * AtMega328p, HC-05 Bluetooth Module, Pulse Oximeter and Heartrate sensor, HX711 weighing Load cell, MLX90614ESF Infrared non-contact temperature sensor.   **Accenture, Hyderabad 02/2017-10/2017**  ***Associate Software Engineer***   * Worked on Data analysis and planning. * Got trained in Python, Data structures, C#, MSSQL, ASP.NET & Microsoft SharePoint. * Reporting of the error in PO/ PR to the Electrolux – client and raised tickets on the behalf of clients against the issue and resolved them in ServiceNow on priority basis.   ***Project***  **Electrolux – JDE PR/PO**   * Support and development of the robust web application, JDE-PR/PO, to help, identify and surface the content that client consider to be the most useful and relevant, via the **Analytics Processing Component** in SharePoint Server to analyz both the purchase orders of APAC region itself, and also the way that users interact with the listings by fetching data from MSSQL servers. |