SAMANVITHA BURLE

Fairfax, VA 22030 | P: +15714884823|
E: **samanvithaburle1997@gmail.com|**
https://www.linkedin.com/in/samanvitha-burle
https://github.com/Sburle43

PROFESSIONAL SUMMARY

- Skilled Data Analyst with 4 years of experience in Data Analysis, Machine Learning, and Data Mining of structured and unstructured datasets.
- I hold a graduate degree with a commendable 3.83 CGPA, and with a proficient background in Data Acquisition, Validation, Predictive Modeling, Data Visualization, and Web Scraping.
- Certified IBM Cloud Pak for Data Solution Architect with a strong command of Statistical Programming Languages such as R and Python.
- Expert in Text Analytics, Statistical Machine Learning, and Data Mining solutions for diverse business problems.
- Proficient in developing Data Visualizations using R and Python.
- Skilled in multiple cloud services such as IBM Cloud Pak for Data, Microsoft Azure, and Google Cloud Platform.
- Experienced in creating AUTO AI models using Microsoft Azure Auto AI, IBM Auto AI, and H2o.
- Successfully designed and executed the full AI lifecycle on Account Receivables utilizing IBM Cloud Pak for Data.
- Proficient in managing diverse data science projects such as Sales Forecasting, Customer Classification, Survival Analysis, Sentiment Analysis, Text Mining, and Recommendation Systems.
- Expertise in transforming business requirements into analytical models, designing algorithms, building models, and developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data.
- Proficient in performing data parsing, manipulation, and preparation utilizing techniques such as descriptive data content analysis, computation of descriptive statistics, regular expressions, data splitting and combining, remapping, merging, subsetting, reindexing, melting, and reshaping.
- Extensive experience in utilizing various packages in R and Python such as ggplot2, caret, dplyr, weka, models, twitter, NLP, scikit-learn, pandas, NumPy, Seaborn, SciPy, Matplotlib, sci-kit-learn, Beautiful Soup, and Selenium to perform data analysis, statistical modeling, and data visualization.
- Demonstrated expertise in extracting and processing images and text using Python libraries such as Tesseract OCR, OpenCV, Pillow, scikit-learn, nltk, Beautiful Soup, and Selenium.
- Proven record of creating visually appealing and informative dashboards utilizing tools such as ggplot, Matplotlib, Seaborn, Tableau, R Shiny, Power BI, and Apache Superset
- Experience in building deep learning models such as SSD, YOLO, utilizing frameworks such as TensorFlow and Keras.
- Strong understanding of deep learning concepts such as deep neural networks, artificial neural networks, convolutional neural networks, and recurrent neural networks.
- Proficient in writing complex SQL queries including stored procedures, triggers, joins, and subqueries.
- Thorough understanding of Hadoop Architecture and its various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, MapReduce concepts, and ecosystems including Hive and Pig.
- Experienced in Python data manipulation for loading and extraction as well as utilizing Python libraries such as NumPy, SciPy, Pandas, and Spark 2.0 (PySpark, MLlib) to develop various models and algorithms for analysis.
- Proficient in Test-Driven Development and Agile-Scrum Development methodologies
- Proficient in using GIT Version Control System.
- Effective team player with strong leadership skills, the ability to work independently, excellent communication and interpersonal skills, highly organized with efficient time management skills, and the ability to handle multiple tasks and thrive in a team environment.

TECHNICAL SKILLS

Big Data	Hadoop, PySpark, Hive, HDFS
Languages	Python, R, SQL, Nodejs, JavaScript
ETL Tools	TensorFlow, PySpark
Databases	MySQL, Mongo DB, PostgreSQL,
Data Visualization Tools and Libraries	MS Excel, Tableau, Power BI, ggplot, seaborn, matplotlib
Machine Learning Algorithms	Logistic Regression, Linear Regression, Support Vector Machines, Naïve Bayes, Random Forest, Decision Trees, K-Nearest Neighbors, Gradient Boost, K-Means Clustering

Deep Learning Algorithms	Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, Computer Vision Techniques, Yolo, SSD
Data Analysis Libraries	Pandas, NumPy, SciPy, Scikit-learn, Statsmodels, NLTK, Matplotlib, OpenCV, Tesseract OCR, Python Flask, Keras, seaborn and ggplot.
Cloud Computing Tools and Frameworks	Google Cloud Platform, Databricks, IBM Cloud Pak for Data, IBM Watson Studio, Apache Superset, Microsoft Custom Vision, Microsoft QnA Maker, Microsoft Azure Auto AI, Visual Studio, and Anaconda
Platforms	Windows, Mac, Unix/Linux

WORK EXPERIENCE

George Mason University - Research Assistantship **August 2022 – May 2023**

- Performed advanced image and text analysis research using a wide range of tools and methodologies, including machine learning, computer vision, and Python programming language.
- Conducted comprehensive data analysis to identify patterns and trends in image and text data, and recommended areas for further research.
- Developed custom Python web scraping scripts to collect data from Glassdoor, NNDB, and other websites that were not easily accessible by existing tools.
- Implemented automated data extraction and cleaning processes using Python libraries to increase efficiency and analyze SEC Edgar filings.
- Developed a web crawler using Python and Selenium to collect data from the NNDB website and created a dataset for further analysis.
- Generated a web crawler using Python and Selenium to download the top Glassdoor reviews for various companies.
- Conducted extensive data scraping from Edgar filings using Python libraries such as Requests and Beautiful Soup and analyzed the image characteristics.
- Implemented an image captioning model using BLIP2 to generate captions for images from Edgar filings.
- Developed a zero-shot image classification model to categorize images from Edgar filings.
- Employed PyTorch and pre-trained CLIP models to predict the aesthetic quality of images scraped from Edgar filings.
  Contributed to the development of research proposals and grant applications by providing expertise in image and text analysis methodologies.
  Technical Stack: Python, Microsoft Excel, Selenium, Pandas

Spruce Technologies, Inc. – Backend Developer **August 2022 – December2022**

- Developed a SQL-based relational database to support the backend functionality of a web application.
- Created an optimized data model that facilitated efficient data storage, retrieval, and management within the web application.
- Collaborated with the team to implement new features such as user authentication and authorization, leveraging the data model's capabilities.
- Designed and implemented a web crawler using Python libraries like Pandas, NumPy, and Beautiful Soup to automatically identify the functioning of URL links.
- Developed an automated Python program using regular expression functions, Pandas, and NumPy libraries to replace broken URLs with working ones, enhancing website usability.
- Validated the implementation and verified the proper functionality of URLs using a Python script.
  Technical Stack: JavaScript, HTML, Python, Pandas, SQL

Cognizance Technologies, LLC. – Data Analyst **May 2022 – August 2022**

- Collaborated closely with the front-end development team to comprehend the data model requirements and guarantee smooth integration between the front-end and back-end systems.
- Devised a comprehensive data model for website transactions and purchases utilizing PostgreSQL.
- Implemented database schemas, including tables, views, and stored procedures, to support the data model.
- Created and executed SQL queries to extract data from databases and perform data analysis.
- Designed and established RESTful APIs for the application using Nodejs, allowing the front-end to interact with the back-end data model seamlessly.
- Designed and produced business reports using Apache Superset.
- Presented reports to a diverse range of stakeholders such as Administration, ISR, and Executive, to communicate key findings and insights effectively.
  Technical Stack: Node Js, JavaScript, Apache Superset

Miracle Software Systems, Inc – Data Analyst **April 2018 – July 2021**

- Leveraged IBM Cloud Pak for Data to develop a solution that categorizes customers with similar account receivables characteristics, providing account managers with timely monitoring and enhancing customer services.
- Utilized Data Virtualization to establish connections with multiple databases, allowing virtualization of data for streamlined analysis.
- Implemented data governance rules with IBM Cloud Pak for Data to safeguard customer details.
- Utilized IBM Auto AI to create a 90% accurate classification model for categorizing customers with comparable account receivables characteristics.
- Created reports using IBM Cognos Dashboard and presented them to multiple stakeholders for account receivables management.
- Developed an automated process for categorizing returned envelopes, reducing the company's workload.
- Utilized OpenCV and Pillow in Python to preprocess images of envelopes and extract information.
- Conducted text analysis using Python and Tesseract OCR to extract the required details such as name, date, and reason for return from the envelopes.
- Built a business bot using Microsoft QnA Maker to extract question-answer pairs from user manuals and policies, improving the user experience by providing faster access to information.
- Proficient in developing AUTO AI models using various cloud-based platforms such as Microsoft Azure, IBM Cloud, and H20.
- Developed a Python-based classification model using Microsoft Custom Vision with 85% accuracy to categorize receipts and bills from various companies, streamlining the organization process for seamless refunds.
- Created an Android app using Python Flask and HTML, with a machine learning model built in Python and Microsoft Custom Vision, for identifying animal and tree species from uploaded images.

Technical Stack: Python, IBM Cloud, Microsoft Custom Vision, IBM Watson Studio, Visual Studio Code, TensorFlow

PROJECTS

Detection of Chest X ray abnormalities using DICOM Images:

- Developed a computer vision model that classifies and localizes lung abnormalities from x-ray images.
- Used DICOM x-ray images of human chest areas to train the model and automate the detection process.
- Implemented an automated boundary-drawing feature to pinpoint the area of abnormality and facilitate rapid case analysis.
- Employed best practices in machine learning and computer vision to ensure high accuracy and reliability of the model.
- Collaborated with healthcare company to ensure that the model meets their needs and can be integrated into existing clinical workflows.

Technical Stack: Python, Open CV, TensorFlow, Computer Vision

Detecting Abnormal Human Behavior:

- Developed a Python model utilizing OpenCV and dlib libraries to detect abnormal human behavior within an organization.
- Trained the model on video data and utilized computer vision techniques to accurately identify potentially hazardous behavior.
- Designed and implemented an alert system that immediately notifies security personnel in the event of an abnormality.
- Collaborated with the client to ensure the model meets their needs and integrated it into the organization's safety and security protocols.

Technical Stack: HTML, Python, Python, OpenCV

EDUCATION

George Mason University, M.S in Data Analytics Engineering **GPA** – 3.83/4 **August 2021 – August 2023**

Course work: Capstone Project, Big Data Needs, Business Analytics using R and Introduction to NLP

Andhra University, B. Tech in Computer Science **August 2014 – April 2018**

Course work: Design and Analysis of algorithms, Data Structures, Object Oriented Programming, Software Engineering

CERTIFICATIONS

- IBM Cloud Pak for Data Solution Architect V2.5
- IBM Cloud Pak for Data V3.0.x Data Science
- Google Data Analytics Professional Certificate