

LAKSHMI VENKATA SAILESH VEMULA

Austin, TX, 78753 | +1(317)-371-4658 | vlvsailesh21@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

TECHNICAL SKILLS

- Python, R programming, Html, CSS, PHP, C, JavaScript, PySpark, MATLAB, MySQL, PostgreSQL, Elasticsearch, HBase, Hive, Pig, Oozie, Sqoop, MongoDB, NoSQL, Pandas, Matplotlib, NumPy, scikit-learn, SDLC, Agile, Power BI, Tableau, SSIS, GitHub, MS Excel, Visual Studio

EXPERIENCE

DATA ENGINEER | INDIANA BUSINESS RESEARCH CENTER 07/2020 – 12/2020

- Writing python scripts for automating to fetch and integrate data into SQL servers, API from the web.
- Reviewed, Debugged, and improved the performance of R, Python, PySpark scripts, and created clear technical documentation.
- Engaged in creating a Microsoft SQL Master Data Management System that will provide the demographic data of all the patients affected by the COVI-19 all over the world scraped from various trusted resources. For doing this ETL pipelines are created SSIS packages that can extract, clean, and load data automatically.
- Experienced in debugging, manipulating data, cleaning, analyzing, and extraction from web servers with a vast range of python libraries like NumPy, Pandas, BeautifulSoup, Matplotlib, Seaborn, ggplot.
- Developed various dashboards using Tableau having data points coming from various resources like flat files, MySQL databases, and Hadoop File systems.
- Helped in creating a Master Data Management System using Hadoop Ecosystem technologies namely, PIG, HIVE, and HBase. Sqoop is used for integrating preexisting SQL databases and Oozie is used to create Job schedulers.
- Developed real-time data pipelines to extract streaming data using Apache Kafka and Spark Streaming.

DATA ENGINEER | POLIS CENTER 01/2020 – 07/2020

- Modeled and Developed a MySQL-based Data Warehouse and created a python script to get the data from NoSQL Database (MongoDB) based API service.
- R programming-based scripts are written to get the data from API services and aggregating them in one place. Gathered data is then analyzed for various Key Performance Indicators of a business.
- Engaged in developing various machine learning algorithms such as linear regression, Random Forests regression, K-means clustering, KNN classification, and various other supervised learning algorithms using python libraries like Pandas, NumPy, SciPy, NLTK, TensorFlow, Scikit-learn, PyTorch, and Keras.

RESEARCH ASSISTANT | INDIANA UNIVERSITY PURDUE UNIVERSITY INDIANAPOLIS 08/2019 – 01/2020

- Developed a website based which will facilitate students to enroll in the program offered by the client as well as will provide help from start to end of their program, there is also an admin dashboard where the client can see analytics based on the data provided by the students (like discussions, grades). For developing this website from scratch technologies such as CSS, PHP, Html, JavaScript, SQL, are used.
- Developed a web interface that runs on python as a scripting language to run a machine learning algorithm in the backend using the Django framework.

DATA SCIENTIST | TATA CONSULTANCY SERVICES – INDIA 09/2017 – 07/2019

- Worked on Business Intelligence projects of analyzing data from banks and providing suitable inputs to the bank.
- Participated in the full product development life cycle including analysis, design, development, deployment, and operations support.
- Utilizing Web scraping techniques to extract and organize suitable vendors information using Selenium, Beautifulsoup packages.
- Built and tuned multiple machine learning models for finding best suitable resource in company for given requirements.
- Various types of data and big data which are in HDFS are connected to Tableau and created interactive Dashboards.
- Based on the Analytics provided by Tableau, KPIs are found. The SparkML library is then used in the PySpark to predict the performance based on the KPIs given.
- Developed complex jobs to perform CRUD operations and analysis of data in Cassandra distributed database using Spark DataFrame API.
- Created Oozie workflows that collect data from the live SQL servers and load them into HDFS using Sqoop.
- Created different Dashboards and reports using Microsoft Power BI and shared them over the client network, so that they can do presentations without having any local file copy which decreases the hassle of maintaining files.

PROJECTS

Gender Classification and Age Estimation using Convolutional Neural Networks Sep 2020 – Dec 2020

- Developed a Deep Learning Model that used more than 5 GB of image dataset to predict the person's gender and age group and annotate over the image or live webcam feed or recorded video.
- Worked on Deep Learning Libraries (TensorFlow, Keras), Graphical Processing Unit, Computer Vision Library (OpenCV), and Jupyter Notebook for developing the model.

Data Visualization of US Foreign Trade Statistics Sep 2020 – Dec 2020

- This project is about analyzing the US census trade data which contains all the imports and exports related information of all states in the USA. Particularly, it shows the different commodities that are traded, and the value (In millions of dollars) of the trades.
- Html, CSS, JavaScript, and d3.js were used to develop the dashboard.

Object Detection and anonymizing Oct 2020 – Dec 2020

- Developed a custom-built Deep Convolutional Neural Network model to detect the person wearing a cap and hiding the identity of only that unique person wearing a cap. To read, process, and anonymize images Computer Vision Library (OpenCV) was used.

Predictive Analysis Using restaurant Zomato data Jan 2020 – Apr 2020

- Managing and analyzing a large data of 1 GB using Spark RDD's transformation, Spark SQL. Performed feature selection to figure out the optimal feature set.
- Building a Model to predict how much overall rating a restaurant will get if it starts a cuisine and the estimated amount of cost for two people. Developing a sentiment analysis of reviews given by customers using NLP which tells whether it is a positive or negative review.

Dashboard of World population statistics Jan 2020 – Apr 2020

- Explored different flat files containing hundreds of rows, cleaned, and transformed according to my specification, and presented in a clear understandable report using Tableau tool.

EDUCATION

- Master of Science | Applied Data Science** GPA: 3.9/4
Indiana University Purdue University Indianapolis Aug 2019 – Dec 2020
- Bachelor of Technology | Electronics and Communication Engineering** GPA: 7.8/10
Jawaharlal Nehru Technological University Kakinada Aug 2013 – May 2017