**SUMMARY:**

Overall, 10 years of IT Experience as a Data Science Developer and Data/Software Engineer and with over 7 years in Python, SQL, AWS and ETL. A Problem Solver and Prolific Researcher with good Data Engineering and Analytical skills and experience of working in all stages of Data Lifecycle.

**SKILLS:**

* Programming: Python, AWS, MySQL, TensorFlow, DynamoDB, Airflow, Kubernetes, Docker, Azure Shell Scripting.
* Technologies: Machine learning (ML), Deep Learning (DL), Databases, Data Lake, NLP, OCR, Computer Vision, Tesseract, git, Linux.

**EMPLOYMENT EXPERIENCE:**

**LexisNexis, Sacramento-CA Apr 2022 – Present**

**Software Engineer (Data),**

* Helped Design and Build EHR Data Processing System completely on Cloud (AWS, Azure) and Load Resulting Structured Data onto Data Lake. It involves Python, MySQL, DynamoDB, S3, Machine Learning, NLP, Computer Vision and Shell Scripting.
* As part of Data Engineering effort Built ETL pipelines to clean, transform and normalize Named Entity information using Python and Pandas. We used DynamoDB as a central database to orchestrate all the pipelines.
* Helped Design and Build SQL Schema to store all the resultant Structured Data onto the Database and join with other existing health Databases.
* Worked on a Huge Data Analysis Exercise to understand the nature of the Raw EHR Data and Explored various Solutions for each of the components in the system either existing services or building custom solutions.
* Used Existing Document AI Services (Computer Vision) and built custom models on top of pre-built models to PDF Documents to text format.
* Used Existing Medical Text AI Services (NLP) to extract Named Entity information from Unstructured text documents. Example: Drug Names, Disease Diagnosis, Lab Reports.
* Worked with Business Analysts to gather and analyze requirements and come up with technical solution to solve it. Create Agile sprint stories by dividing the solution into manageable parts and identify any upstream or downstream dependencies.
* Technologies used are Python, Pandas, NumPy, AWS, EC2, MySQL, S3, ETL, Testing, NLP, ML, Linux and Shell Scripting.

**LabCorp, Raleigh-NC Jan 2018 – Feb 2022**

**Data Science Developer,**

* As part of Data Engineering effort built an Automated workflow in AWS to Generate Label Data from existing Data using Python, S3, DynamoDB, Lambda, tesseract.
* As part of Data Engineering effort-built Workflows for ETL, pre/post processing of the labelled data using Python, AWS, SQL and store the data into the Databases.
* Helped Design & Built Continuous Learning System where it Automatically retrains the model periodically on failure cases using TensorFlow extended (Airflow, TensorFlow, AWS).
* Worked on a Data project to Extract, Transform and Load EHR information from Patient Records using
* Python, AWS, SQL and load all the Data onto the Data Lake as part of the Analytics requirements.
* Built a DL Model to Classify Payer and Predict Member I’d location on the Insurance Card using Python, TensorFlow, AWS.
* Built a DL Model to Predict Front/Back of the Insurance card and another model to Predict Orientation using Python, TensorFlow, AWS.
* Built a Utility Models on Smaller scale using human labelled dataset and used this model to create more training data.
* Built Data Cleaning Pipelines by Analysing label data Anomalies using Data Visualization tools like Matplotlib.
* Built a Smart Cropping System, using a custom written Algorithm developed from Heuristics by Analysing hundreds of scenarios.
* Worked with Business Analysts to gather and analyse requirements and come up with technical solution to solve it. Create Agile sprint stories by dividing the solution into manageable parts and identify any upstream or downstream dependencies.
* Technologies used are Python, Pandas, NumPy, AWS, EC2, DynamoDB, S3, Airflow, MySQL ETL, TensorFlow, TFX, Data Engineering and Shell Scripting, Linux.

**Software Engineer (Data), LexisNexis, Sacramento-CA Mar 2017 – Nov 2017**

* Worked on an Automation Project which Collects the Legal documents, Converts and Enriches the data completely in cloud, which was previously done by humans thereby saving millions of dollars every year.
* Built an Attachment Removal System, using Heuristics developed from analysing hundreds of documents.
* I majorly worked on Document Classification, Auto-Conversion, ETL, Image Processing, OCR and Enrichment of the data using Machine Learning, Python and AWS and store all the Data onto Data Lake.
* Built a ML model to analyse actual text from Noise in a document Image by creating 10 custom features (Feature Engineering) by analysing the nature of the documents.
* Worked with Product Manager to gather and analyse ML requirements and come up with technical solution. Helped Create Agile sprint stories by dividing the solution into manageable parts and identify any upstream or downstream dependencies.
* Technologies used are Python, MySQL, ETL, scikit-learn, scikit-image, Olena (Image Processing), Tesseract, AWS, Shell Scripting, Linux, Data Mining and git.

**Center for Digital Humanities –University of South Carolina, Apr 2015 – Dec 2016**

**Developer,**

* Worked on a Major Academic digital project for SCIAA. Job includes Extraction (OCR) of text from Archaeological site forms and creating a database for all the site forms (50000+) of South Carolina state.
* Built a Noise cleaning Algorithm, using Heuristics developed from analysing hundreds of documents.
* Created a Document Searching website, which allows researchers to search across all the materials contributed by the Federation members. Technologies used are Solr, Ruby on Rails, Java, MySQL, Docker and Git.
* Worked with Humanities Professors to gather and analyse their requirements and come up with technical solution. Come up with a plan by dividing the solution into manageable parts and create a lean tech stack.
* Job involves OCR (Tesseract), Image Processing (MATLAB), Machine Learning (WEKA) and Text Processing (Python), R.

**Syntel, Chennai -India Dec 2012 –Dec 2014**

**Software Engineer,**

* Our team built and maintained an ETL System, Which Extracts credit card transaction data, transforms it and loads it into hundreds of databases.
* Converting the existing Mainframe system to Ab Initio, uses Ab Initio, Shell Scripting and DB2 and Java.
* Job Responsibilities include coding and testing (Unit, Assembly & Regression).

**EDUCATION:**

* **Master of Engineering in Computer Science (3.4) Dec 2016,**

University of South Carolina – Columbia

* **Bachelor of Technology in Information Technology (7.1) May 2012,**

Gokaraju Rangaraju Institute of Engineering and Technology, JNTU, Hyderabad, INDIA