Sudharsan Annamalai

Sivanthangal, Chennai.

+91-7904383847
sudharshanvas@gmail.com
LinkedIn Profile

Technical Skills and Interests

Languages: Python, Java, SQL, Linux comands
Developer Tools: Vs code, PyCharm, Eclipse, jupyter notebooks(Google-colab)
Software tools: Microsoft excel, Microsoft Powerpoint
Developer Tools: Vs code, PyCharm, Eclipse, jupyter notebooks(Google-colab)
Frameworks: Selenium, Pandas, Tkinter, Numpy, scikitlearn, Tensorflow
Cloud/Databases: AWS, Amazon Quicksight, PowerBI, DynamoDB, BigQuery, DataProc, GCP
OS: : MacOS, Linux
Areas of Interest: Data Engineering, Data Science

Certifications

–Google IT Automation with Python
Google (Link)

–Deep Learning specialization
DeepLearning.AI (Link)

–SAS Visual Business Analytics
SAS (Link)

–Introduction to Data Visualization using Google Data Studio
Coursera Project Network (Link)

Experience

•Amazon *Jul 21*
Risk Management Coordinator Chennai

- Working on process queries and complains specific to copyright risks and violations associated with Kindle digital books.
- Assisted in the data collection and analysis process to consolidate parameters related to potential risky customers.
- Assisted in the development and enhancement of automated tools and systems to monitor and identify potential copyright infringements, allowing for proactive detection and mitigation of risks.
- Automated the data collection process using Selenium to scale down the process time of our workflow.
- Served as the dedicated BI champion within the team, assisting in writing SQL queries to meet diverse process requirements from the workforce.
- As a BI, I have assisted in creating visualization for Customer Value analysis, Risk score Analysis, and Workflow management tasks using AWS QuickSight.
- Led the onboarding and training of 16 new hires, equipping them with essential workforce process knowledge, SQL, and advanced Excel techniques with PowerQuery for effective data management.

•The Institute of Mathematical Sciences (IMSc) *Jan 19 - Feb 20*
Computer Programmer Chennai

- Implemented an OCR pipeline with Tesseract (an opensource OCR engine) to extract the Latin characters, and chemical names from the PDFs scanned images from literatures, and the compendium of Indian Medicinal plants.
- I have created and implemented an ETL pipeline for collecting the data of essential oils in Indian Medicinal plants in Google Cloud Platform
- Assisted in the extracting the tabular data from PDF files with tabula-py and Apache Spark(pyspark).
- I have assisted in creating BigQuery tables from different, and also helpe creating numerous queries to pull data from BigQuery for downstream reasearch analysis.
- Contributed to the development of a comprehensive database by organizing and structuring collected data, ensuring data integrity and accessibility for further analysis.
- Conducted data validation and verification using pandas and bash automation scripts, identifying and resolving any discrepancies or errors to maintain data accuracy.
- Supported the team in data analysis tasks, including generating reports, visualizing data using PowerBI, and Googel Charts.
- Developed and implemented an automated process for Chemical ID Mapping using Selenium framework and Python.

•Scope e-Knowledge Center *Jan 16 - Feb 18*
Content Moderator Chennai

- Worked on moderating the content in medical website Medscape, and a related application called "Consult".
- Volunteered and assisted in creating automation tool to detect HIPAA violations from text and images.

Academic Projects

•Machine learning model(Leprosy) *2016*

Designing and implementing a machine learning model to predict hyper-active reactions in leprosy patients

- Tools & technologies used: python, scikit-learn,GraphPad Prism 6 (Statistical tool)
- Descriptive analytics and the basical level interpretations about the causation of the hyper active reactions were understood using GraphPad Prism
- python and pandas was used for the data preparation and data validation tasks.
- Our prototype and model was developed with scikit-learn Decision tree algorithm.
- CART DecisionTree Analysis engine was used to build the Decision tree, and the generated model was validated with 10-fold cross-validation method.

•Prediction of secretory profiles from fungal genome *2018*

Designing and implementation of a data pipeline

- Tools & technologies used: Linux Scripting, python, BigQuery, GCP
- Designed and implemented a data pipeline to predict the secretory signatures from fungal pathogens using genomic data
- This project has helped us to understand the potential drug targets in the spectrum of fungal secretome.

•Bio curation and screening of toxic compounds *2018*

Curation and Database development

- Tools & technologies used: Linux Scripting, python, BeautifulSoup, SQL
- Compiled and created a database specific to essential oil through curation of Ayurveda-based research articles and books, and traditional databases
- implemented python automations and bash scripting to create a ETL pipeline.
- This project has helped us to understand the potential drug targets in the spectrum of fungal secretome.

PUBLICATIONS

Chaitanya, V. Sundeep, et al. "Analysis of a novel multiplex polymerase chain reaction assay as a sensitive tool for the diagnosis of indeterminate and tuberculoid forms of leprosy." International journal of mycobacteriology 6.1 (2017)

Education

•X-Standard *2010*

Adhiyaman metriculation Higher Secondary school, Uthangarai. Percentage: 86.00

•XII-Standard *2012*

Vidya Vikas metriculation Higher Secondary school, Tiruchengode. Percentage: 77.50

•BTech Biotechnology *2016*

Sree Sastha Institute of Technology, Chennai CGPA: 6.27

•MTech Biotechnology *2019*

Sree Sastha Institute of Technology, Chennai CGPA: 8.17