santokalayil@gmail.com
+91 8891960880
Singasandra, Bengaluru
www.santokalayil.com

SANTO K THOMAS

Data Scientist / Python Developer

SUMMARY

Highly analytical and process-oriented data scientist, with in depth knowledge of database types, research methodologies, machine learning models, highly adept at accurate curation, analysis, and visualization of data. Possess extensive analytical skills, strong attention to detail, team player, resilient with a strong work ethic to succeed. Dream to initialize or to be part of at least one innovation in the field of Data Science for this millennium.

- ◆ Github: https://github.com/santokalayil
- ◆ LinkedIn: https://www.linkedin.com/in/santokalayil
- ◆ Kaggle: https://www.kaggle.com/santokalayil

SELECTED PROJECTS

- HAWK: a Social Media Analysis and Fraud Detection tool: developed as part of a global hackathon conducted by RBI on the topic 'Social Media Analysis for detection of digital payment fraud and disruption'.
- Churn Analysis and Modelling Project: Getting data from Amazon Redshift using SQL (via MODE), processing, cleaning and verifying the integrity of the data, getting features out of data, selecting features, building, optimizing ML models, finding out features that cause churn and suggest measures to reduce churn
- Capstone Project: Cancer (Biopsy) prediction with 95.7 testing accuracy on patient survey data. Tested multiple Classification methods and selected ML models were based on Logistic Regression Gradient Boost Classifier.
- Employee Attrition & Performance Analytics: Achieved 60% faster detection of attributes that determine and influence employee performance and attrition so that the HR department can introduce new cost-effective strategies to improve productivity and performance
- Bank Personal Loan Modelling Data Science Project - Designed an optimal and efficient model that could help a large bank identify potential customers who have a higher probability of purchasing loans which will increase the success ratio while at the same time reduce the cost of campaigns. Built ML models with Specificity up to 96%.
- Automated Directory Project: Developed a directory of community members by automating processes such as data collection, database storage, data retrieval and analytics, rearranging data for various sections of the book, data correction, Image resizing, full PDF generation etc. Technologies and Tools used are Google Forms, Google Sheet API, Python (Pandas, Numpy, Flask, Matplotlib, Reportlab, Pillow, OpenCV, Sqlite etc).

KEY SKILLS

- Machine Learning: Supervised, Unsupervised Learning (Regression, Classification, PCA, Clustering.) and Computer Vision (OpenCV) and basic Deep Learning.
- Data Analysis & Visualization: using Python, Pandas, Numpy, Matplotlib, Seaborn, Plotly, Plotly Dash,Tableu
- Interpreting ML Results using LIME & SHAP interpretation techniques
- Statistics, Problem Solving
- Programming - Python (Advanced), Rust (Basic), Julia(Basic). Python based-skills in automated pdf report and document generation using Report Lab Python and Social Media Automation. GUI App Creation using FLTK Rust, Kivy Python
- Database: MySQL, Postgres, Sqlite, MongoDB
- Operating Systems: Linux (Bash & Python Scripting), Windows. Command Line Skills (Linux terminal, Windows Command Prompt and Powershell)
- Data Collection using Web-scraping and Web-automation tools like Beautiful Soup, Selenium, Requests, PyAutoGUI
- Experience with Docker, Git, web-APIs, Insomnia, Postman, OpenCV, NLTK, DASK, Gradio, Streamlit
- Coordinate, communicate and provide technical support to other functional groups relating to web applications.
- Other skills: HTML5, CSS, Bootstrap 4, Tailwind CSS, Django, Flask, Google Forms, Google sheet API, Adobe InDesign, Affinity Publisher, Microsoft Office, etc.

PROFESSIONAL SUMMARY

- Consultant Data Scientist at Ezetap Payment Solutions (November 2021 - present day): Worked on Churn Analysis & Prediction project (getting data from Amazon Redshift using SQL (via MODE), processing, cleaning and verifying the integrity of the data, getting features out of data, selecting features, building, optimizing ML models, finding out features that cause churn and suggest measures to reduce churn. Currently working on a computer vision project (using OpenCV) to detect and classify objects and PII (personal identity information) in images from social media posts.
- Freelance: Data Collection and Automation, Analytics and Directory DTP Book Generation for Hennur MCC, Bangalore

- Website Developer & Administrator at Sacred Heart Matriculation School, Padi, Chennai (until 2019)
- Freelance: Web Developer & Administrator for Diocese of Pathanamthitta, Carmal Central School, Carmala, Pathanamthitta, Kerala

EDUCATION

Course	Institution	Year	Remarks
Postgraduate Programme in Data Science and Engineering	Great Lakes Institute of Management	2020	Certificate Link: https://verify.greatlearning.in/FFEJOUXA
B Th.	Urbaniana University, ROMA	2014	85%
B A-Philosophy	University of Kerala	2010	62%
Higher Secondary	St. Mary's Higher Secondary School, Pattom, Trivandrum	2007	82%
SSLC-10	Sacred Heart High School, Mylapra, Pathanamthitta	2004	80%

OTHER CERTIFICATIONS & ACHIEVEMENTS

- Runners Up in HARBINGER 2021, a global hackathon conducted by RBI on the developing a tool on the topic "Social Media Analysis for detection of digital payment fraud and disruption"
- ABRSM- Royal School of Music, London - Certified in Music Theory
- Author of MALLU python package for Easier Machine Learning Workflow- published under MIT License in The Python Package Index (PyPI) Software Repository. Available via "pip install"