Kanak **Pandit**

DATA ENGINEER · PRODUCT DEVELOPER

Allmandring 16 A 28, Vaihingen, Stuttgart 70569, Germany

 +49 162 7776170 | st194112@stud.uni-stuttgart.de | kanakpandit17.github.io/Portfolio/ | kanakpandit17 | /kanak-pandit-92740214b/ |

Kanak Pandit

"Code the future with intelligence."

Summary

Goal-driven professional with over 3 years of experience in cloud computing, data engineering, and full-stack development. Skilled in data science, analytics, and artificial intelligence with a strong command of SQL, Python, C/C++, JavaScript, AWS, Power BI, Looker, Mage, and Spark. Proven track record of building secure and scalable cloud solutions, designing and implementing ETL processes, and applying machine learning algorithms for data-driven decision-making. Expertise in developing dynamic web applications and delivering solutions that optimize business processes. Eager to contribute to innovative projects that leverage cutting-edge technology and deliver impactful outcomes.

Work Experience

Nirmaan Hyperloop
Mumbai, India

SOFTWARE ENGINEER AND RESEARCH LEAD   ·   *Feb. 2021 - July. 2024*

- Developed and implemented control systems for the Hyperloop pod, focusing on the GUI using PyEdifice.
- Integrated the system with Phoenix Contact's PLCnext for seamless communication and control.
- Conducted comprehensive analysis of social media discussions using LDA topic modeling, sentiment analysis, sarcasm detection, and opinion inversion techniques to evaluate public sentiment towards hyperloop technology. Identified key areas of concern using ML Algorithms and sentiment labelling techniques like BERT and VADER.

Tech Mahindra
Mumbai, India

DATA ENGINEERING(AWS) INTERN   ·   *Jun. 2023 - Sep. 2023*

- Led the design and implementation of a scalable video processing system using AWS services, enhancing user experience with secure S3 storage, automated transcoding, metadata extraction, and dynamic ad insertion.
- Implemented user authentication and leveraged AWS Rekognition and MediaTailor for efficient content delivery.

RAA Technocrates
Mumbai, India

SOFTWARE DEVELOPER INTERN   ·   *Dec. 2022 - Mar. 2023*

- Documented software requirements, defined project scope, and ensured alignment with objectives.
- Developed a dynamic educational web application from scratch using MongoDB, AngularJS, and TailwindCSS, providing a user-friendly interface and tailored content for students.

Projects

Streamlined Data Processing with No-code ETL

- Developed a No-code ETL tool to simplify data management, centralize diverse data sources, automate processing, and ensure data quality, empowering non- technical users and enhancing data-driven decision-making.
- Implemented functions for converting data between formats such as CSV and JSON, using libraries like Pandas for efficient data handling. Validate conversions through comprehensive testing to ensure data integrity, including handling edge cases.
- Developed algorithms to extract data metadata, including missing values, data types, and unique values, using Pandas or custom solutions. Conduct extensive testing with diverse datasets to ensure accuracy and scalability.
- Created visualizations (e.g., pie charts, bar graphs) using libraries like Matplotlib, Seaborn, or Plotly. Test the visualizations for accuracy and relevance by comparing with known data patterns across different datasets.
- Implemented mechanisms for ingesting real-time data streams using hashing.

Data Engineering ETL Project on Uber Expenses Tracking

- Designed and implemented a modern data pipeline using Google Cloud Platform (GCP) tools like Storage, Compute Instances, BigQuery, and Looker Studio to process and analyze large-scale Uber trip records, ensuring efficient data ingestion, transformation, and reporting.
- Developed Python scripts for data preprocessing and analysis, leveraging GCP services to clean, aggregate, and transform the TLC trip record data into actionable insights, enabling improved decision-making and business analysis.
- Utilized Mage Data Pipeline Tool for orchestrating ETL processes and automating data workflows, contributing to the open-source community by actively contributing to the Mage AI GitHub repository.
- Created interactive dashboards using Looker Studio to visualize key metrics from Uber data, such as trip distances, fares, and passenger counts, providing stakeholders with real-time business insights and performance tracking.

Sentiment Analysis of Hyperloop as Future Transportation System

- Identified key areas of public discussion on hyperloop, aiding in the categorization of concerns and interests within the technology's development.
- Classified social media comments into positive, negative, and neutral categories, revealing public sentiment trends and influencer engagement.
- Analyzed sarcastic remarks and opinion contradictions in retweets, shedding light on message consistency and digital marketing implications.
- Provided actionable insights for financial models and marketing strategies by understanding public discourse patterns, contributing to hyperloop's future promotion and adoption.

Dynamic Pricing System for Hotel Booking Management

- Developed SmartStay, enabling comprehensive guest, room, and reservation management with integrated Dynamic Pricing strategies to optimize revenue.
- Used Tkinter for GUI, MySQL Workbench with MySQL Connector, PyMySQL, and SQLAlchemy for database interactions, and Plotly/Matplotlib for data visualizations.
- Integrated Selenium for capturing graph screenshots, FPDF for generating PDF reports, and Yagmail for automated email dispatch to guests and hotel administrators.
- Utilized Google Generative AI (Gemini) for generating plot descriptions, providing valuable insights in analytics reports for hotel management.

Market Basket Analysis

- Developed and optimized association rule mining algorithms using Python libraries like Mlxtend, including implementations of Apriori, FP-Growth, and ECLAT, enhancing data mining efficiency for industry decision-making.
- Applied data analysis and visualization techniques using libraries such as NumPy, Pandas, Matplotlib, Seaborn, and Plotly, creating interactive and informative data insights for stakeholders.
- Explored improvements in association rule mining through statistical analysis of lift and confidence, proposing advanced methods like time-series clustering to enhance pattern recognition and recommendation systems.

Data Analysis and Visualization of Brazilian E-commerce Market using Olist Dataset

- Performed comprehensive data wrangling and exploratory data analysis for e-commerce datasets, handling missing values, duplicates, outliers, and performing normalization to prepare high-quality data for analysis.
- Utilized data visualization tools like Streamlit to create an interactive sales performance dashboard, featuring filters for customer segmentation and KPIs, providing actionable insights for business stakeholders.
- Designed and analyzed relational data models using Entity Relationship Diagrams (ERD) to capture customer, order, payment, and product relationships in a comprehensive e-commerce business schema.
- Conducted in-depth analysis of customer purchase behavior, on-time delivery rates, and product performance through RFM segmentation and other visualization techniques, driving data-driven decision-making.

Honors & Awards

INTERNATIONAL

2022	**1st Prize,** MULTICON-W Engineering Workshop and Paper Presentation Competition	*Mumbai, India*

DOMESTIC

2021	**Amongst 2 percent Toppers,** Python for Data Science organized by NPTEL, IIT Madras	*Mumbai, India*
2022	**Amongst Top 12 Teams,** Smart India Hackathon at Institute Level	*Mumbai, India*
2021	**Amongst Top 5 Teams,** Google Developer Student Club in Android App Development	*Mumbai, India*

Education

University of Stuttgart — *Stuttgart, Germany*

MSc IN COMPUTER SCIENCE — *Oct 2024 - Present*
- Chosen courses in Artificial Intelligence, Data Integration, Data Warehousing and Mining, Cloud Computing, Information Visualization

Thakur College of Engineering and Technology — *Mumbai, India*

BSc IN COMPUTER SCIENCE — *Jun 2020 - Jun 2024*
- Chosen courses in Software Development and Architecture, Artificial Intelligence and Machine Learning, Web Scrapping, Data Engineering, Cloud Computing