DataHack 2.0 Abstracts & Problem Statements

1.Data Science (Powered by Streamlit):

Problem Statement:

The Indian startup ecosystem has experienced remarkable growth and funding from 2018 to 2023, particularly in sectors like Artificial Intelligence (AI) and Electric Vehicles (EV). However, a comprehensive analysis of the market trends and underlying factors driving this growth is needed. This problem statement aims to address the challenges of conducting a detailed market analysis, identifying growth factors, deriving data-driven insights, and developing an interactive dashboard using Streamlit. By addressing these challenges, participants will contribute to a better understanding of the Indian startup ecosystem's growth in the AI and EV sectors, providing valuable insights for entrepreneurs, investors, policymakers, and other stakeholders.You can use openly available news and/or any publicly available datasets for performing this case study.

Brownie Points

1. Teams will receive bonus points for utilizing comprehensive and relevant datasets, demonstrating data relevance, extent, and coverage.
2. Extra recognition will be given to teams that create an enhanced and interactive experience for potential investors through engaging and interactive dashboards.
3. Teams that develop and apply their own unique evaluation metrics will earn bonus points, showcasing their creativity and advanced analysis skills.
4. Additional recognition will be awarded to teams that accurately predict startup growth based on the available data, demonstrating their ability to analyze trends and make informed predictions.

General Questions you can answer

1. What are the top 10 startups in terms of revenue over the last 5 years;
2. Give a location-wise distribution of the top 100 startups;
3. Give a comparison between Foreign and domestic funds and the sectors they are investing in;
4. Valuation based analysis on Fintech companies over the last 5 years;
5. How did Covid affect funding for startups across different sectors;
6. Which sectors have had the most number of new and emerging startups;
7. Compare the growth rate of start-ups with their age ;
8. How do funding rounds affect the growth of startups;
9. How has funding for different sectors, like EV evolved over time;
10. Startups in which sectors have been stagnating or decreasing;

2. MLOps (Powered by LawYantra)

Abstract:

In a rapidly evolving legal landscape, the demand for effective client-attorney matchmaking has never been higher. To meet this challenge, we invite participants to develop an innovative MLOps-driven platform that seamlessly connects clients with legal experts. This platform will harness the power of Machine Learning Operations (MLOps) to facilitate precise matches based on client's unique legal concerns and attorney's specialized expertise. Our goal is to revolutionize the legal industry by optimizing the client-attorney matchmaking process

Problem Statement:

MLOps-Powered Legal Matchmaking: Create an MLOps-driven platform for precise client-attorney matchmaking, leveraging advanced techniques to connect clients with attorneys based on their unique legal concerns, while ensuring that the platform can accept user requests. This datathon aims to streamline legal expertise access, ensuring a user-friendly and efficient experience, ultimately revolutionizing the legal industry. The platform should be capable of extracting necessary information from user requests in any language for efficient matchmaking. For example, match a client seeking patent advice with an attorney specialized in intellectual property law, enhancing the accessibility and effectiveness of legal support.

Dataset :
Following is a real world data of Lawyers
https://docs.google.com/spreadsheets/d/15Okg9hJvMsOqBXWEzZXSmX30T0pXK6NckoEuIkiYnbU/edit?usp=sharing

Key Objective:

1. Client-Attorney Precision Matchmaking: Create a recommendation system to connect clients with attorneys who match their specific legal needs. System should also ensure that recommended lawyers should have an equal gender ratio.
2. Efficiency and User-Friendly Experience: Optimize matchmaking and simple query input for a seamless experience.
3. Data Storage: Create a vector database to store and search the word embeddings of the lawyer data.
4. Client-Server Interface: Create a Client-Server Interface where users can put their queries and get recommendations.

Brownie points:

- Feedback Integration: Integrate a feedback system for clients and attorneys to provide reviews and ratings, and display this to the admin.
- Data Preprocessing: Working with real world data comes with its own set of problems, preprocessing data and removing anomalies is important.
- Explainable AI Component: Develop a feature that explains why specific attorney-client matches are made, enhancing transparency and trust.

- Multilingual Support: Implement the capability for the platform to support multiple languages, making it accessible to a global audience. This task involves language detection, translation, and providing attorney-client matching across language barriers.
- AI Query Assistance: Integrate real-time AI support for user queries, enhancing precision and effectiveness in attorney-client matchmaking, especially for non-native languages.
- Analytics Dashboard: Extract Data from the given dataset to create an analytics dashboard.

- Evaluation criteria:
- You will receive the test data at 9am on 29th October, 2023 the judging round, which will include information about more lawyers. Using the admin dashboard, you should be able to add these new profiles to the system. The recommendation feature should also dynamically include these newly added profiles in real-time, ensuring that they are part of the recommendations made to users.

3. Artificial Intelligence

Abstract:

In a world increasingly focused on health and fitness, the need for a tech-savvy companion that can accurately track and guide one is more vital than ever. One major part of this companion is the presence of a pose estimation algorithm for both tracking and guiding one in an activity. This should be seamlessly integrated with the health statistics of the user to create a holistic fitness experience for the user.

Problem Statement:
Traditional fitness tracking methods often rely on manual input, which can be imprecise and lead to suboptimal results, impacting the user's overall fitness journey.

Your task here is to explore and develop a state-of-the art pose estimation system using Computer Vision and Machine Learning, to be used in the field of sports and fitness. You can choose the following use case or choose your own use cases.

Gym Companion:-

- Exercise Recognition:

Users often struggle with exercise recognition and selection. Manual entry of exercises is error-prone and can disrupt the flow of a workout. Leveraging OpenCV and deep learning, this system offers automated exercise recognition. Users can seamlessly perform a variety of exercises, such as Bench press, Pushup, Bicep Curl, Lateral Raise, Shoulder Press, and Squat, without the need for manual input.

- Rep and Set Counting:

Inaccurate rep and set counting can hinder the effectiveness of workouts, leading to under-training or over-training. Using OpenCV, the system sets angle thresholds, enabling accurate counting of repetitions. It also tracks the number of sets completed, ensuring users follow their workout plans with precision.

- Progress Tracking:
Fitness enthusiasts often lack a centralized platform to comprehensively track and visualize their fitness progress over time. The system not only collects and stores critical user data, encompassing sets, reps, weight lifted, BMI, and heart rate, but it also incorporates a data analysis dashboard. This user-friendly dashboard allows individuals to monitor their progress, gain insights from the collected data, and identify areas for improvement.

Additional Brownie Points:
For extra recognition, consider including the following features or aspects in the project:

- Exercise Recommendations:Users may need guidance on exercise selection and workout optimization. The system utilizes collected data to offer exercise recommendations tailored to users' fitness goals, helping them diversify their routines and achieve better results.
- Warning system: The system includes a valuable safety feature – a warning system. This system not only tracks your progress but also closely monitors your exercise form. If your form begins to deteriorate to a point that may risk injury or hinder the effectiveness of your workout, the warning system provides immediate alerts, helping you maintain a safe and efficient workout routine.

- App Integration: The system can be integrated into a mobile app, offering real-time feedback and data accessibility across devices, potentially connecting with wearables

4.NLP (Powered By LawYantra)

ABSTRACT:

In today’s digital world, the need for enhanced accessibility to legal cases has become of utmost importance. Legal documents in various languages often hinder efficient comprehension and retrieval. This project seeks to develop a state-of-the-art portal using NLP and LLM technology to overcome language barriers, enabling the translation of legal case PDFs into English and their secure storage

PROBLEM STATEMENT:

The objective is to create a portal for legal case documents that is capable of translating legal case PDFs from languages such as Spanish, Hindi into English using state-of-the-art NLP(Natural Language Processing)/LLM(Large Language Model) technology. Additionally, store these translated PDFs. When users have inquiries or need specific information related to legal cases, the portal should swiftly and accurately respond with the most pertinent data extracted from the translated PDFs using word embedding stored in a vector database. This not only enhances access to legal content but also ensures that responses to user queries are highly reliable. Participants will receive 20 PDFs containing legal cases from Lawyantra.

FEATURES:

- Translation Capabilities: Develop a robust translation system using NLP and LLM technology to accurately convert legal case PDFs from languages like Spanish into English.
- Transliteration Capabilities: Create a robust transliteration system using NLP and LLM technology to accurately convert legal phrases from languages like Spanish into the English script.
- User Query Handling: Create an intuitive search interface that allows users to ask questions related to the legal case documents.
- Contextual Retrieval: Ensure that the portal can understand and retrieve contextually relevant information from the translated PDFs in response to user queries.
- Accuracy: Prioritize accuracy in translation and retrieval to provide precise and dependable responses to user questions.

BROWNIE POINTS

Create a Client Server Interface for the users to query PDFs.

EVALUATION CRITERIA :

The evaluation will be done by the output generated by queries provided by Lawyantra.

5.NLP (Powered by Fueler)

Abstract:
In today's digital landscape, the ability to demonstrate one's expertise holds exceptional value, and "proof of work" is the cornerstone of this demonstration. This concept encapsulates the practice of showcasing tangible evidence of one's skills, knowledge, and accomplishments.
It is a dynamic and ever-evolving testament to your capabilities in the digital age. It enhances your credibility and trustworthiness, opens doors to career opportunities, and fuels your motivation to keep learning and improving . A Digital portfolio helps one to seamlessly work with like-minded professionals and thereby inspiring growth and success

Problem Statement:

The challenge at hand is to create an AI-powered system capable of automatically recognizing and categorizing projects uploaded within a portfolio. This system should be capable of identifying various content types within the projects, such as text, images, charts, and other visual elements. Additionally, it should analyze the portfolio data to provide personalized project suggestions to users, thereby increasing their chances of securing opportunities like internships, jobs, or freelancing work. For example, the AI system has proactively recognized the user's desire to expand their skills into AI. It not only suggests AI-related project ideas but also offers comprehensive support, guidance, and resources to facilitate the user's journey into the AI domain. This personalized approach helps the user diversify their portfolio and, ultimately, enhances their opportunities in the competitive job market.

Main Features:

- The system should be able to parse uploaded resumes in pdf / other formats and thereby provide recommendations on improvement.
- The system should be able to tag and organize the projects automatically, making it easier for users to manage their portfolio.
- The system should be user-friendly and easy to use.
- System should recommend people of similar interests/projects in a user’s recommendation.
- Implement features to provide tools for tracking project progress and performance.

Brownie Points :

- Collaborative Workspace:
Create a collaborative platform where users can invite collaborators to work on projects together in real-time, making it a hub for team-based projects and collaborative learning.
- Multi-Lingual Support
Incorporate language diversity by supporting multiple languages, making the platform accessible to a global user base and facilitating multilingual project descriptions and content recognition.

- Mobile Application

Develop a mobile application for users to manage their portfolios on the go, enhancing accessibility and allowing them to capture and add projects from their mobile devices.

- Advanced Analytics and Engagement Dashboard:

Offer an advanced analytics dashboard with data visualization tools, enabling users to gain deep insights into their project performance and audience engagement. Offer support to boost users' opportunities by increasing visibility, engagement, and relevance of their portfolio projects

JUDGING CRITERIA :
Projects will be judged on how accurately your product will recognize and categorize the content. Also based on ease of navigation and user friendliness. Innovation and creativity of your product and its features and how relevant its predictions are. The product should be able to handle different file formats and types of data across different languages. The analytics provided should be helpful and insightful.