**Applied Data Science - Predicting IMDB scores**

**PHASE I**

**Problem Definition:**

* **Problem Statement:**

The problem at hand is to build a predictive model that can accurately forecast IMDb scores for movies and TV shows based on various features and attributes associated with the content. This problem falls within the domain of applied data science and machine learning.

* **Problem Scope:**

Understand the dataset through statistical analysis and visualizations, identifying patterns and correlations in the data.

* **Problem Goals**:

The primary goal of this applied data science project is to develop an accurate and reliable predictive model for IMDb scores of movies and TV shows. By leveraging machine learning algorithms and advanced data analysis techniques, the aim is to provide viewers, filmmakers, and streaming platforms with a tool that can forecast IMDb ratings based on various features associated with the content.

**Design Thinking Steps:**

* 1. **Empathize:**
* Understand Stakeholders: Identify and empathize with stakeholders such as viewers, filmmakers, and streaming platforms. Understand their perspectives, goals, and challenges related to IMDb scores.
* User Interviews and Surveys: Conduct interviews and surveys to gather insights about what viewers and industry professionals value in a movie or TV show.
* Data Gathering: Collect historical IMDb scores, user reviews, and metadata to understand existing patterns.
* 2. **Define**:
* Problem Definition: Clearly define the problem of predicting IMDb scores. Consider the specific challenges faced by users and the industry, such as uncertainty in audience reception and decision-making processes.
* Identify Constraints: Understand technical, budgetary, and data-related constraints that might impact the solution design.
* Create User Personas: Develop user personas representing different types of users, their goals, and their interactions with IMDb scores.
* 3. **Ideate**:
* Brainstorming Sessions: Conduct brainstorming sessions with a multidisciplinary team to generate a wide range of ideas for predicting IMDb scores.
* Ideation Workshops: Organize workshops to encourage creativity and innovative thinking among team members.
* Crazy Eights: Use techniques like Crazy Eights to prompt participants to sketch eight ideas in five minutes, encouraging rapid ideation.
* 4. **Prototype**:
* Create Prototypes: Develop prototypes of the IMDb score prediction system. This could include a basic model using a small dataset to demonstrate the concept.
* User Feedback: Gather feedback from stakeholders and users by demonstrating the prototype. Use their input to refine the prototype further.
* Iterative Prototyping: Iterate on the prototype based on user feedback, refining the user interface and prediction algorithms.
* 5. **Test**:
* User Testing: Conduct user testing sessions where stakeholders interact with the prototype. Observe how they use the system and gather feedback on its functionality and usability.
* A/B Testing: If applicable, conduct A/B testing with different versions of the prediction system to identify the most effective solution.
* Iterative Refinement: Continuously refine the system based on user testing feedback, aiming to enhance user satisfaction and the accuracy of IMDb score predictions.
* 6**. Implement**:
* Full-Scale Development: Develop the complete IMDb score prediction system based on the refined prototype.
* Integration: Integrate the prediction system into relevant platforms, such as streaming services or movie databases.
* User Training: If necessary, provide training sessions or documentation to educate users on how to effectively use the IMDb score predictions for decision-making.
* 7**. Evaluate**:
* Performance Metrics: Evaluate the system’s performance using metrics like accuracy, user engagement, and user satisfaction.
* Feedback Analysis: Analyze user feedback and make further improvements to the system based on real-world usage.
* Impact Assessment: Assess the impact of the IMDb score predictions on user decision-making processes, content creation strategies, and viewer satisfaction.
* 8. **Iterate**:
* Continuous Improvement: Continuously gather user feedback and iterate on the system to adapt to changing user needs, technological advancements, and industry trends.
* Incorporate New Data: Integrate new data sources and machine learning techniques to enhance the accuracy and relevance of IMDb score predictions over time.
* By following these design thinking steps, the applied data science project can ensure that the IMDb score prediction system is not only technically robust but also user-centered, addressing the unique needs and challenges of stakeholders in the entertainment industry.

**Conclusion:**

In conclusion, the journey of applying data science techniques to predict IMDb scores has been a transformative process, bridging the gap between raw data and actionable insights for the entertainment industry. Through a meticulous approach that combined data exploration, algorithmic innovation, and user-centric design, this project has successfully provided valuable solutions and profound insights to viewers, filmmakers, and streaming platforms alike.