Ben Ryan Rinconada BSIT-3A

ITBAN2 - Knowledge Discovery

01. Think of examples where you can apply knowledge discovery. Give at least five

- **Network Intrusion Detection:**
  The area of application of Knowledge Discovery would be in the analysis of network traffic data; knowledge discovery will allow the identification of patterns indicative of normal versus malicious activity. Machine learning algorithms could be applied to detect anomalies that might be indicative of network intrusions. This could entail predictive models that forecast prospective threats based on historic data for preemptive network protection. It applies real-time data mining techniques that permit constant monitoring of network traffic for patterns, which are abnormal, to immediately detect and act upon intrusions.
- **University Course Recommendation System**
  Application of Knowledge Discovery: Analyzing the past academic records would help in discovering patterns of student performance and in recommending courses that match their strengths and interests. Discovery of patterns in students' interests and course content provides personalized course recommendations that match their career goals and preferences. Discovery of the at-risk students through enrollment and academic performance patterns allows interventions to improve the retention rate.
- **Medical Diagnosis Support**
  Application of Knowledge Discovery: Patient data, such as symptoms, medical history, and lab results, are analyzed to predict the likelihood of diseases or conditions in general so that early diagnosis and intervention can be made. Identification of the most effective treatment is achieved by discovering patterns and correlations in patient outcomes with respect to the different treatments. Generating models suggesting personalized treatment plans for a patient based on their unique data can help obtain better health results.
- **Social Media Analysis**
  Knowledge Discovery Application: User posts and comments are analyzed to understand public sentiment toward a brand, product, or service. Companies can then make sense of customer opinions and use them for enhancing offerings. It can also discover new trends and topics from the hashtags, keywords, and engagement of users to guide marketing and creation. Understanding the behavior of users and their engagement patterns can help to fine-tune content delivery for maximum retention of users and satisfaction.
- **Website Clickstream Analysis**
  Application of Knowledge Discovery: The analysis of clickstream data in finding out how users navigate across a website may help to discover common paths and bottlenecks, which helps in the improvement of site design and user experience. Segmenting users by their browsing behavior provides targeted content and advertisements, enhancing engagement and conversion rates. Knowing factors that contribute to successful conversions by examining sequences of clicks leading to purchases or other desired actions helps to optimize marketing strategies and website layout.

02. Think of an organization in which you can use CRISP DM or KDD. Which of the two process models do you think is more appropriate to use? Explain why.

Organization: University

For a university aiming to implement a Course Recommendation System, the CRISP-DM (Cross Industry Standard Process for Data Mining) model would be more appropriate than the KDD (Knowledge Discovery in Databases) process model. Here's why:

Structured Framework:

CRISP-DM provides a well-defined, structured approach with six distinct phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. This structure helps the university systematically address each aspect of the recommendation system development process, ensuring thorough analysis and effective implementation.

Business Understanding:

The initial phase of CRISP-DM focuses on understanding the business objectives and requirements. In the context of a university, this means clearly defining the goals of the course recommendation system, such as improving student satisfaction, aligning course offerings with career goals, and increasing retention rates. This phase ensures that the data mining efforts are aligned with the university's strategic objectives.

Iterative and Flexible:

CRISP-DM is inherently iterative, allowing the university to revisit and refine previous steps based on new insights. For example, after developing an initial model, the university can evaluate its effectiveness and make adjustments to improve accuracy. This flexibility is crucial in an educational setting where requirements and data can evolve over time.

Comprehensive Evaluation:

The Evaluation phase of CRISP-DM ensures that the developed models meet the desired objectives and perform well with the actual data. This phase is critical for the university to validate the effectiveness of the course recommendations and make any necessary adjustments before full deployment.

Deployment and Monitoring:

CRISP-DM includes a Deployment phase, where the recommendations are put into use, and a monitoring system is set up to track performance and gather feedback. This is essential for the university to ensure the recommendation system remains effective and can be continuously improved based on real-world usage and feedback from students and faculty.

03. As a future IT professional which part of the Kimball Lifecycle could you be potentially involved in? Explain Why

Being a prospective IT professional, I can engage in several parts of the Kimball Lifecycle, particularly in:

Data Staging and ETL: Data from different sources is collected here, cleaned, and transformed into a format ready for analysis. As a part of ETL processes, my role could include ensuring that the data is clean and, in a form, fit for the task, a critical point in the process for effective reporting and analysis.

Data Modeling: The data warehouse schema is designed, including fact and dimension tables. This will include collaborating with business users to understand the requirements and translate them into a robust data model.

Data Presentation and Reporting: Reports, dashboards, and data visualization tools are prepared that help end-users make data-driven decisions. It includes understanding business needs and creating intuitive and user-friendly interfaces for data access.