Based on the **GTU Paper Format** and the specific **Course Outcomes (CO5 & CO6)** associated with this unit, **Unit 5: Tools and Applications** is the bridge between theory and industry.

In a 70-mark paper, this unit typically accounts for **12–14 marks**, appearing primarily in **Question 4** and **Question 5**. While the questions are often descriptive, they carry high weightage for **Understanding (U)** and **Application (A)** cognitive levels.

Unit 5: Predicted Question Bank (Tools & Applications)

1. Most Repeated / High-Probability Questions

Following the (a) 3-mark, (b) 4-mark, and (c) 7-mark structure of your model paper.

[Short Answer Type - 03 Marks]

- **List any three popular Data Mining tools** used in the IT industry (e.g., Weka, Orange, RapidMiner).
- **Define Web Mining.** Mention its three main sub-types.
- **What is Sentiment Analysis?** Give one real-life example of its use.
- **Define Spatial Data.** How does it differ from traditional relational data?

[Descriptive Type - 04 Marks]

- **Explain Web Content Mining vs. Web Usage Mining.** Provide a clear distinction with examples.
- **Describe the features of the Weka tool.** Why is it preferred for academic and diploma projects?
- **Explain the application of Data Mining in the Retail Industry** (specifically Market Basket Analysis).
- **What is Text Mining?** Briefly explain the process of extracting patterns from unstructured text.

[Long Answer / Diagram-Based - 07 Marks]

- **Discuss the applications of Data Mining in Banking and Finance.** Focus on **Fraud Detection** and **Credit Scoring** systems.
- **Explain the types of Web Mining in detail.** Draw a neat hierarchy diagram showing Content, Structure, and Usage mining.
- **Describe the Data Mining application in Healthcare.** How can it help in disease prediction and hospital management?
- **Compare Commercial Data Mining tools vs. Open Source tools.** Provide at least five

points of comparison including cost, support, and flexibility.

---

2. Application & Logical Thinking Questions

These questions differentiate high-scoring students by testing system-level understanding.

21. **System Interpretation:** A telecom company is losing 10% of its customers every month. Which data mining application should they implement to identify these customers before they leave? Explain the concept of **Customer Churn Analysis**.
22. **Logical Reasoning:** An online news portal wants to show "Related News" to its readers based on the links they click. Which specific type of Web Mining—**Usage or Structure**—is most relevant here? Justify your answer.
23. **Problem Solving:** You are given a dataset of 5,000 customer reviews for a new smartphone. Explain the steps you would take using **Text Mining** to categorize these reviews as "Positive" or "Negative."
24. **Real-life Application:** How can **Spatial Data Mining** be used by a city's traffic department to reduce congestion during peak hours? Mention the type of data (GPS/Maps) involved.
25. **Tool Selection:** If you are working for a startup with zero budget for software, which category of tools (Open Source or Commercial) will you choose? Name one specific tool and explain one challenge you might face.

Exam Success Strategy for Unit 5:

- **The "Application" Focus:** If the question asks for applications in Banking or Retail, don't just write one line. Use sub-headings like *Risk Management, Fraud Detection, and Customer Profiling*.
- **Web Mining Diagram:** This is a very common 7-mark question. Ensure you can draw the tree structure for the three types of Web Mining.
- **Tool Familiarity:** Since this maps to **CO5**, if you have used **Weka** or **Orange** in your practicals, mention specific features like "The KnowledgeFlow interface" or "Pre-processing filters" in your theory answers to impress the examiner.

Next Step: This concludes our unit-wise Predicted Question Banks! Would you like me to generate a **Full-Length 70-mark Mock Exam Paper** following the exact GTU format and CO/Cognitive level distribution?