

Predicting High-Probability Exam Topics

Problem Statement

Identifying and predicting potential exam topics is a challenge faced by students and educators. Although question papers often repeat similar patterns, most analysis remains manual and time-consuming. The goal of this project is to build a **machine learning-based system** that analyzes past years' exam papers and predicts **high-probability topics** for future exams. The system aims to output a ranked list of high-probability topics for upcoming exams, enabling students to prioritize their preparation effectively.

Existing Solutions and Gaps

1. ExamPredict (<https://exampredict.com>)

What it does: An AI-powered platform that predicts “high-probability questions” using data from previous exams.

Strengths: Detects recurring questions and presents results through a simple web interface.

Gaps: It focuses mainly on **question prediction**, not **topic forecasting**, and is a **closed-source** commercial product with no model transparency or interpretability.

2. Predicting Difficulty of MCQs (PapersWithCode)

What it does: Predicts the difficulty of multiple-choice questions using linguistic and semantic features.

Strengths: Demonstrates how ML and NLP can analyze textual questions.

Gaps: Focuses on **difficulty estimation**, not **topic trends**, and depends on NLP pipelines — not ideal for numerical, trend-based datasets like past paper topics.

3. Automatic Question Paper Generation (IJER Journal)

What it does: Uses keyword extraction and ML algorithms to generate question papers automatically from a syllabus database.

Strengths: Ensures topic coverage and balanced question distribution.

Gaps: Operates on **static question banks** and lacks **time-series prediction** of topics across years.

Summary of Gaps:

Existing tools mainly target question or paper generation, not **topic prediction**.

None integrate **topic frequency**, **marks weight**, and **yearly trends** within a single transparent model. Hence, there's a clear opportunity for a lightweight, interpretable system dedicated to **topic trend forecasting**.

Our Methods / Approaches

1. Collect past 5–10 years of question papers and preprocess them into a structured dataset.
 2. Use **Logistic Regression** or **Decision Tree** algorithms to predict future topic probabilities.
 3. Apply **trend analysis** to observe recurring topics and weightages over time.
 4. Visualize predictions using graphs for clarity and better understanding.
 5. Evaluate performance using metrics such as **Accuracy**, **Precision**, **Recall**, and **ROC-AUC**.
-

References Taken

- [ExamPredict.com](https://www.exampredict.com/)
- [Predicting Difficulty of MCQs – PapersWithCode](https://paperswithcode.com/paper/predicting-difficulty-of-mcq)
- [Springer: Predicting Topics in Scholarly Papers](https://www.springer.com/9783319555614)
- **ChatGPT** and other online resource

Team Members::

- M. Surya Kapil (CSE - D | Roll Number: 11)
- Khushal Sahu (CSE - D | Roll Number: 50)
- Ramyak Ray (CSE-D | Roll Number: 09)