Project Proposal – PYQ Analyser

Title

An Evidence-Based Exam Preparation Analytics System Using Previous Year Question Papers and Syllabus Alignment

---

1. Introduction & Motivation

Examination preparation in university settings is often constrained by limited time, extensive syllabi, and a lack of clear guidance on topic prioritization. Although students have access to previous year question papers (PYQs) and official syllabus documents, there is no systematic mechanism to analyze these resources together and extract actionable insights.

As a result, students frequently:

- Study the entire syllabus without knowing which topics are most important
- Miss frequently asked or high-weightage topics
- Spend time on low-impact areas
- Experience high stress due to uncertainty before examinations

This project is motivated by the need for a **data-driven, transparent, and exam-aware system** that can analyze PYQs in alignment with the official syllabus and help students make informed preparation decisions.

---

2. Problem Statement

Despite the availability of previous year question papers and detailed syllabi, students lack an intelligent system that can:

- Analyze PYQs automatically
- Align questions with syllabus modules and topics
- Identify frequently asked topics and questions
- Provide module-wise importance and grading trends
- Offer evidence-backed preparation guidance

The absence of such a system forces students to rely on intuition, peer suggestions, or incomplete resources, leading to inefficient preparation and avoidable stress.

---

3. Objectives of the Project

The primary objective of this project is to design and develop an **exam preparation analytics system** that transforms previous year question papers into **structured, syllabus-aligned insights**.

Specifically, the system aims to:

- Analyze multiple PYQs and extract individual questions with metadata
- Map questions to syllabus-defined modules and topics
- Identify topic frequency and repetition patterns
- Detect semantically similar questions asked in different forms
- Analyze question paper styles (descriptive vs numerical)
- Compute module-wise grading distribution
- Provide evidence-backed, step-by-step preparation guidance

---

4. Scope of the Project

To maintain focus and feasibility, the project scope is intentionally constrained:

- Institution-specific (initially VIT-style syllabus and exam patterns)
- Subject-specific (one or two subjects for implementation)
- Exam-aware (CAT-1, CAT-2, FAT)
- Minimum requirement of **at least three PYQs** for analysis
- The system performs **analysis and prioritization**, not question prediction or answer generation

---

5. Key User Questions Addressed

The system is designed to answer the following critical student queries:

1. **What should I study in each module according to PYQ patterns?**
    - Ranked, module-wise topic importance based on historical exams
2. **Which topics are most frequently asked module-wise?**
    - Frequency-based topic ranking with supporting evidence
3. **Which questions are repeatedly asked, even if phrased differently?**
    - Canonical question detection with repetition count and references
4. **What is the average module-wise grading distribution?**
    - Visualized using pie charts and infographics
5. **What is the question paper style?**
    - Analysis of descriptive vs numerical questions with probability graphs

---

6. Data Sources

The system operates using only two authoritative data sources:

- **Previous Year Question Papers**
    - PDF, scanned PDF, image, or document formats
- **Official Syllabus Documents**
    - Module-wise topic definitions

No external textbooks, answer keys, or expert annotations are required.

---

7. System Approach and Methodology

7.1 Data Ingestion

- Accepts multiple PYQ formats (PDFs, images, scanned documents)
- Converts all inputs into machine-readable text
- Extracts questions, sub-questions, marks, year, paper ID, and question number

7.2 Syllabus Parsing

- Extracts total number of modules and topic lists
- Defines the syllabus as a **fixed topic boundary**
- Prevents generation or analysis outside official syllabus content

7.3 Exam-Aware Module Range Selection

- Allows users to select module ranges (e.g., 1–3, 3–5, 1–7)
- Supports CAT-1, CAT-2, and FAT exam structures
- Filters analysis strictly within the selected module range

7.4 Question-to-Topic Mapping

- Uses syllabus-constrained semantic similarity
- Maps each question to one or more syllabus topics
- Assigns confidence scores to mappings
- Explicitly handles cases where questions reference sub-concepts under broader syllabus topics

7.5 Frequency and Pattern Analysis

- Computes topic frequency and repetition trends
- Aggregates marks to derive importance scores
- Identifies canonical (standardized) questions across different phrasings

7.6 Question Style Analysis

- Classifies questions as descriptive or numerical
- Computes distribution and probabilities
- Visualizes results using graphs

7.7 Dependency-Aware Preparation Sequencing

- Detects prerequisite relationships between topics
- Explicitly informs students of required prior knowledge
- Automatically generates a recommended study order when dependencies exist

---

8. Evidence-Based Output Design

A core principle of the project is **explainability and trust**.

For every insight, the system provides:

- Exact question text
- Year of examination
- Question paper ID
- Question number

This ensures:

- No hallucination
- Full transparency
- High student confidence during exam preparation

---

9. Output & Visualization

The system provides a dashboard containing:

- Module-wise topic importance lists
- Most frequently asked topics
- Repeated (canonical) questions with citations
- Average module-wise grading pie charts
- Descriptive vs numerical question probability graphs
- Step-by-step preparation sequences

---

10. Novelty and Contribution

While PYQ repositories and exam preparation platforms exist, this project introduces several novel contributions:

- **Syllabus-constrained analytics** rather than open-ended topic modeling
- **Evidence-first insights** with traceable references
- **Canonical question detection** across varied phrasings
- **Exam-aware module range control**
- **Dependency-aware preparation sequencing**

The novelty lies in the **integration of NLP techniques with educational constraints**, resulting in a practical, explainable exam preparation system.

---

11. Feasibility and Practicality

The project is feasible because:

- It does not require deep subject-domain knowledge bases
- It relies on bounded semantic alignment within the syllabus
- It can be developed incrementally
- Even partial implementation provides meaningful value

The system prioritizes **engineering realism over AI hype**.

---

12. Conclusion

This project proposes an intelligent, evidence-based system that bridges the gap between previous year question papers and syllabus documents. By converting exam history into structured preparation insights, the system empowers students to study smarter, reduce stress, and prepare with confidence.

The project emphasizes transparency, explainability, and academic integrity, making it both practically useful and academically sound.