**Project Topics**

List of Topics

[1. Automated Grading System with inbuilt Plagiarism Checker 2](#_Toc183082941)

[2. Academic Resource Organizer 3](#_Toc183082942)

[3. Student Assistance Chatbot with LLM (Large Learning Models) 6](#_Toc183082945)

[4. Spam Detection in online social networks 7](#_Toc183082947)

[5. Net income transfer feature in a budget app 8](#_Toc183082948)

**1. Automated Grading System with inbuilt Plagiarism Checker**

* **Objective**: Develop a system that can grade short answers or essays automatically.
* **ML Techniques**: Natural Language Processing (NLP) and deep learning models for text analysis.
* **Impact**: Reduces the workload on lecturers by automating routine grading tasks, providing faster feedback to students.

#### **Objective**

The goal of the Automated Grading System is to ease the burden of grading on lecturers by using machine learning to assess student responses—especially for open-ended, short-answer, and essay-style questions. Rather than simply looking for keywords, the system would evaluate grammar, content relevance, logical flow, and even the critical thinking displayed in responses. This approach would provide insightful feedback, particularly for subjective answers, beyond what traditional grading offers. The model will also be integrated with an inbuilt plagiarism checker to ensure that the students; responses are authentic.

#### **ML Techniques**

To implement this, I would use several key machine learning and NLP techniques:

1. **Text Preprocessing and Tokenization**:
   * The system must first clean the input text, removing unnecessary elements (punctuation, excess whitespace) and breaking it down into tokens (words or phrases). Tokenization is essential to prepare text for analysis.
2. **Embedding and Representation Learning**:
   * Techniques like **Word2Vec or BERT embedding** can transform text into vector representations that capture semantic relationships between words and phrases. This allows the system to understand context and meaning, crucial for assessing student answers that might use different wording but convey similar ideas.
3. **Sentence and Document-Level Analysis**:
   * **Transformers** like BERT, RoBERTa, or OpenAI’s GPT series can process larger chunks of text, such as entire sentences or paragraphs. This enables the system to grasp the overall structure, logical flow, and depth of the answer, making it well-suited for essay grading.
4. **Similarity and Relevance Analysis**:
   * **Cosine similarity** or **semantic similarity scoring** can measure how closely a student’s answer aligns with a model answer. By comparing the student's response with predefined high-quality answers, the system can grade based on content relevance.
5. **Sentiment and Grammar Analysis**:
   * NLP models like **GPT-based models** and grammar-checking tools like **Grammarly’s NLP** can assess clarity, grammar, coherence, and even the sentiment or tone of a piece. These aspects are helpful for scoring writing quality in essays. This models can even check the document for if it was constructed by another artificial intelligence program.
6. **Reinforcement Learning for Model Improvement**:
   * Lecturers can provide feedback on the model’s grading accuracy, which can feed into a **reinforcement learning loop**. Over time, the system adapts and improves its grading accuracy based on feedback, becoming a more effective model.

**Impact**

This system has various benefits to the educational sector including:

* **Time-saving**: The model will help to reduce the amount the amount of time spent assessing a student’s work and as such will give the lecturer more time to perform other actions.
* **Reliability**: The model will be reliable in the sense that it will always be consistent in its analyses and its adaptability to multiple situations and responses.
* **Plagiarism checking**: The in-built function of a plagiarism checker will enable the program to search for copyright incidences and as such will make the students research more and make them more authentic in their findings.
* **Unbiased Grading**: The model will be authentic and unbiased making students to get the mark they truly deserve.

**2. Academic Resource Organizer**

* **Objective**: Transform how lecturers and students access educational materials, making the learning experience more efficient, targeted, and impactful.
* **ML Techniques**: NLP for content classification, clustering, and topic modeling.
* **Impact**: Eases the burden on educators by helping them quickly find and organize relevant resources for each subject, aiding lesson planning.

**Objective**

The primary aim of the Academic Resource Organizer is to assist lecturers and students by a**utomatically categorizing** educational resources (e.g., videos, articles, exercises) into specific curriculum topics, **tagging** resources with relevant keywords that make them searchable and filterable, r**ecommending** resources to users based on curriculum needs, student grade levels, learning objectives, and even individual student performance and preferences. An e**xample use case is a** lecturer could input a topic, such as "Artificial Intelligence," and receive organized resources including videos, diagrams, interactive exercises, and articles. Students could also receive personalized recommendations, like beginner-friendly videos or advanced reading based on their progress.

### ****ML Techniques and System Components****

The core ML techniques involved are primarily NLP-focused, given the textual and multimedia nature of educational content. I will use these techniques and fit it into the system.

#### A. **Content Classification**

* **Objective**: Automatically classify resources based on curriculum categories (e.g., Science, Mathematics, Arts) and subcategories (e.g., Computer Science, Biology).
* **ML Models**:
  + **Text Classification** using pre-trained NLP models like BERT, RoBERTa, and custom-trained models, fine-tuned to recognize curriculum categories.
  + For non-text resources (e.g., videos), **Speech-to-Text** can transcribe audio content to text, which can then be classified.
* **Process**:
  + Input data (text, transcribed video/audio) is processed and classified into subjects and curriculum topics.

#### B. **Content Tagging and Metadata Extraction**

* **Objective**: Assign tags to resources to make them easier to search and categorize by keywords.
* **ML Models**:
  + Named Entity Recognition (NER) and keyword extraction algorithms can identify key terms like “climate change,” “photosynthesis,” or “Pythagoras theorem.”
  + **Topic Modeling** can identify key themes in longer documents and generate topic tags.
* **Process**:
  + The model identifies prominent keywords, phrases, and entities in a resource, then creates metadata tags based on these.

#### C. **Resource Clustering for Similarity-Based Grouping**

* **Objective**: Group similar resources together based on topic and complexity level.
* **ML Models**:
  + **Clustering Algorithms** such as k-means or hierarchical clustering to group resources.
  + **BERT Embedding** for semantic similarity—resources with similar embeddings are grouped, so lecturers can see sets of related resources.
* **Process**:
  + Resources are vectorized into high-dimensional embeddings, and similar items are clustered, allowing similar resources to be recommended together.

#### D. **Recommendation System**

* **Objective**: Suggest resources to users based on curriculum needs and personal learning profiles.
* **ML Models**:
  + **Collaborative Filtering** for recommendations based on user interactions (e.g., previously viewed or liked content).
  + **Content-Based Filtering** based on topic, tags, or specific user needs.
  + **Hybrid Models** that combine collaborative and content-based filtering to make recommendations both topic-aware and behavior-aware.
* **Process**:
  + The system learns from user behavior and preferences, adjusting recommendations over time for personalized learning paths.

#### E. **Personalization and Adaptive Learning**

* **Objective**: Provide targeted resources to specific users, adapting to individual student levels or lecturer’s preferences.
* **ML Models**:
  + **User Profiling and Personalization** algorithms that adjust recommendations based on ongoing assessment data, such as quizzes, assignments, or resource engagement.
* **Process**:
  + The system tracks individual progress, engagement metrics, and quiz results, tailoring content recommendations to support specific strengths and address weaknesses.

### ****Impact****

This system has several impactful benefits for education:

* **Efficiency**: Lecturers can quickly find high-quality resources, saving significant time on lesson planning.
* **Quality of Education**: Personalized resources help students engage with material that suits their learning pace and style.
* **Inclusivity**: Supports various learning needs by offering resources at different difficulty levels and formats (e.g., visual, auditory).
* **Scalability**: By automating categorization and recommendations, this system can support large and diverse repositories of resources.

### Student Assistance Chatbot with LLM (Large Learning Models)

### Problem:

**Solution**: This project focuses on creating a customer service chatbot tailored for a specific domain such as a university service desk, or library help desk. Using a smaller, accessible conversational model like lightweight Llama variant you’ll fine tune the model to handle common inquiries specific to the chosen domain. As part of this project, you’ll need to create or collect dataset of customer service queries and responses if one is not readily available. This can involve gathering frequently asked question (FAQs) from relevant sources or creating your own dataset based on typical customer interactions. The chatbot will be deployed with a basic user interface such as command line tool or simple web application, allowing users to ask questions and receive automated responses

### Spam Detection in online social networks

**Problem:** Large amount of fake news and spam information on the internet.

**Solution**: The main aim is to develop a machine learning based detection system that can be used for different online social networks platforms such as Instagram, WhatsApp and Facebook. The research objective will be to implement the model through machine learning in python to confirm the efficiency of the model and build large dataset with more features to be analyzed by the model with the possible side of natural language processing and to improve the accuracy of the model through the use of optimization bio inspired algorithms.

### Excess allowance/ Net income transfer feature in a budget app