**Name:** Nehal Khan

**Title:**

“A Machine Learning-Based Recommender System for Improving Students Learning Experiences.”

**Problem Identification:**

The goal of this project is to recommend appropriate learning strategies to improve students' learning methods for each course. This will help students focus on the topics where they perform poorly. Additionally, the system can provide recommendations for the entire class by calculating the average marks for each assessment and treating it as if it were a single student's performance.

**Challenges or gaps to overcome:**

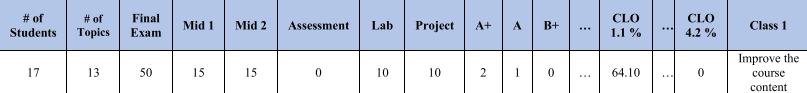
1. **Lack of Systematic Processes**: The existing solution lacks a data-driven approach. For example, if a student performs well in all assessments except one, then the recommendation system should focus on the particular assessment and recommend appropriate actions instead of simply recommending the overall performance as excellent.
2. **Generalizability**: The existing solution did not work well across different courses and curriculum because each course has different strategies and mark distribution.
3. **Limited Dataset**: The existing solution did not take the students’ performance with diverse scoring patterns. Due to this, students who perform well or poorly across all assessments get a valid recommendation, while mixed performances are not.
4. **Ineffective Feature Selection**: Features such as attendance and topics covered per day are irrelevant to the current recommendation system, reducing its effectiveness in identifying important factors affecting student performance.

**Future improvement:**

Integrating teaching strategies and students feedback to improve the recommendations given to the students.

**Inconsistency in Existing Dataset:**

The existing solution for the given problem statement was to give recommendations to the whole class.

****

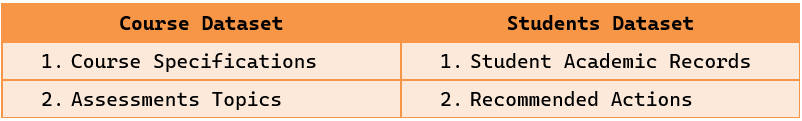
For example, from the existing dataset, it can be seen that the recommendation was focused on just the course content as only 64.10 percent of the course learning outcomes were achieved. The recommendation is also too generic, not specifying what and how to improve it.

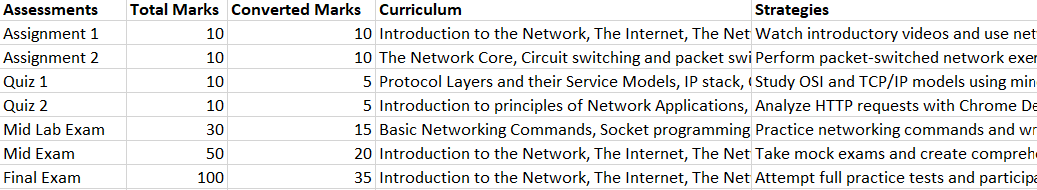
Rather than focusing on what to achieve, the system should focus on how to achieve the remaining 35.90 percent of the course learning outcomes.

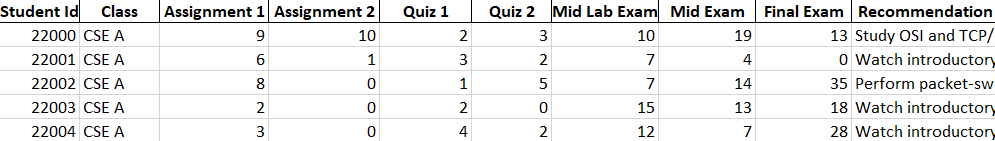
**Proposed Dataset:**

The proposed solution emphasizes generalizability, and to achieve this the dataset had to be generalized in both course specifications and student performance.

The proposed solution includes two separate datasets namely the Course and Students dataset.

****

****

****

From the above figure, it can be seen that the Course dataset contains the course and assessment specifications while the Students dataset contains the academic records of the students.

**Data Preprocessing:**

To make the recommender system generalised, for each course a training dataset is created using the course specifications mentioned in the course dataset.

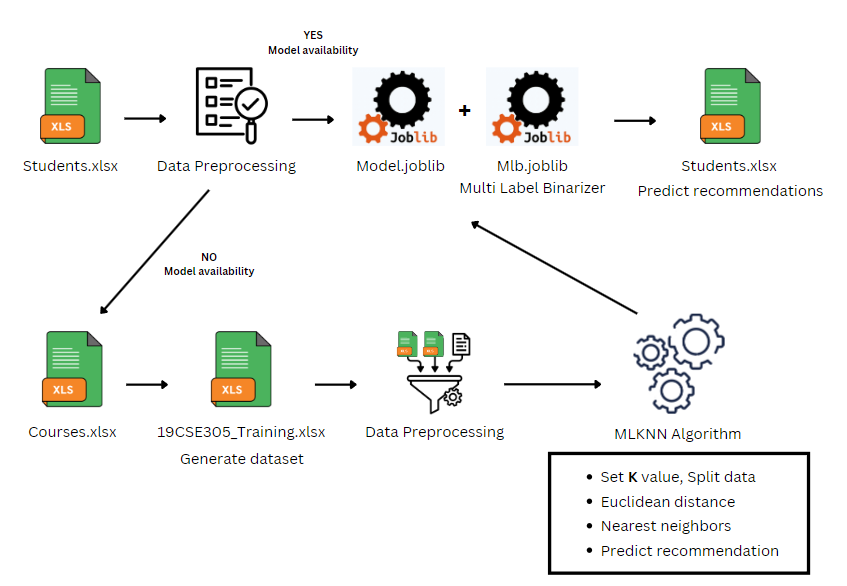
* Build course specification details
* Generate synthetic Data using ‘random.randint()’
* Tutor strategies and student’s performance integration

For the dataset which needs recommendation also undergoes data preprocessing.

* Data cleaning is done to fix missing value
* Convert marks to its ratio as mentioned in course dataset
* Sum the converted marks to get the total

**Problem Approach:**

The method to solve this problem is by implementing a recommender system using the MLKNN algorithm, where each and every strategy recommended by the tutor will be considered as a label. Based on the student’s performance in every assessment, the labels are assigned and finally merged together to give an overall recommendation for the students.

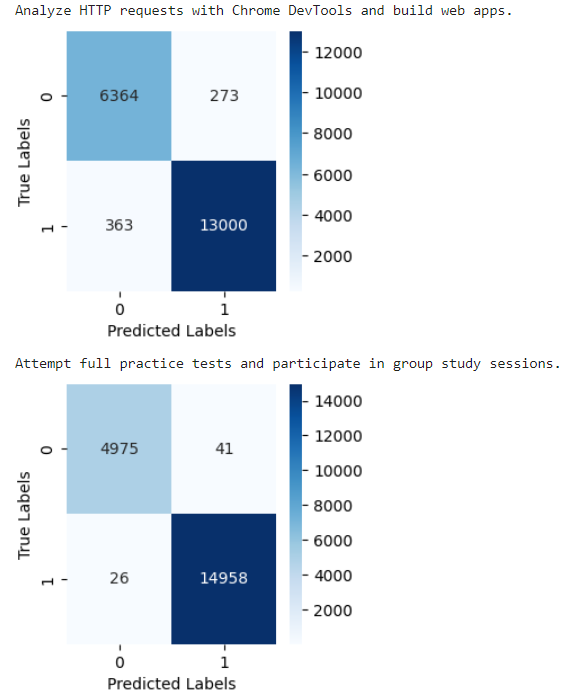
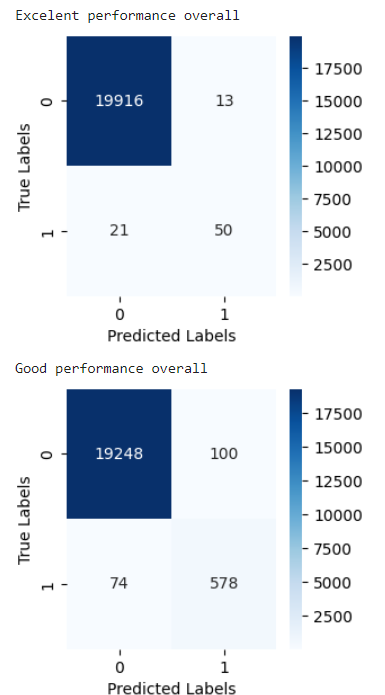


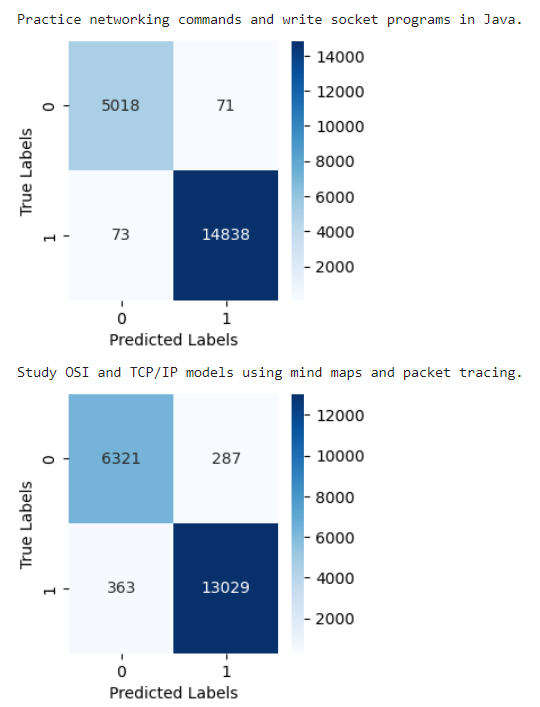
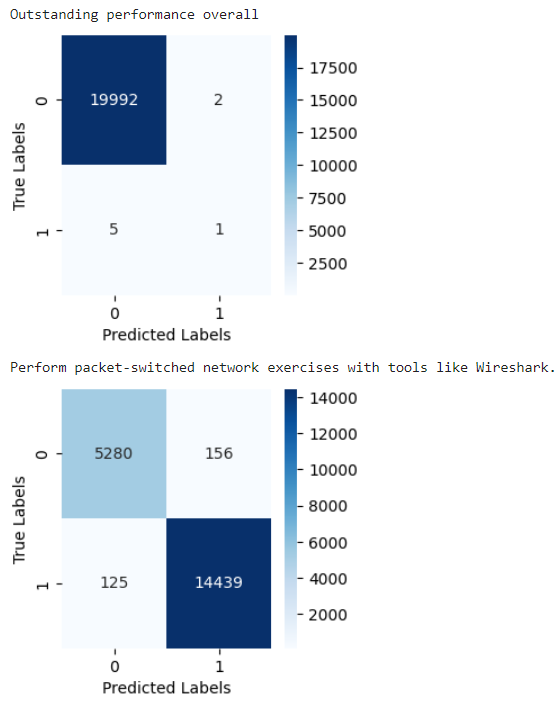
Upon reading the Course dataset the training dataset is generated in the preprocessing step and later data integration is done. Then the final training dataset is passed through the MLKNN algorithm which uses Euclidian distance to calculate the ‘K’ nearest neighbors and predict the most common strategy/recommendation for each assessment, making this a multi-label problem. For the algorithm to understand the recommendations, it converts the recommendations to binary format using a Multi-Label Binarizer. Now using this model, the recommendations for the testing dataset are generated.

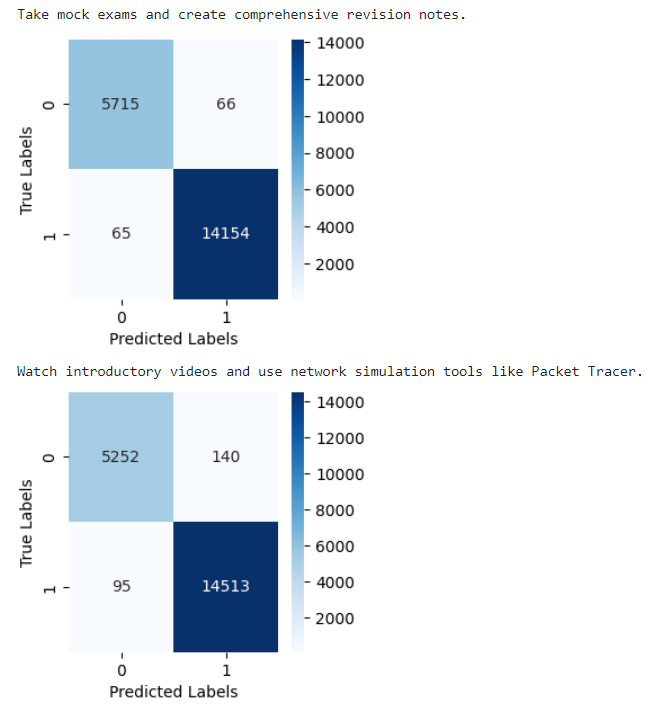
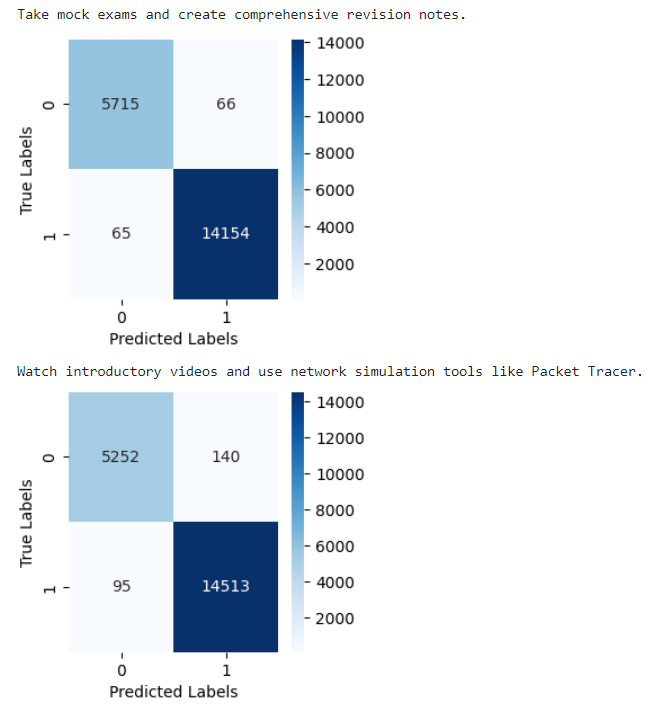
**Results:**

By making the solution generalized, according to the 19CSE301 Course dataset there were total 7 sets of strategies, therefore by including good, excellent, and outstanding performance we get a total of 10 labels, with this as a combination we get over 5040 possible recommendations.

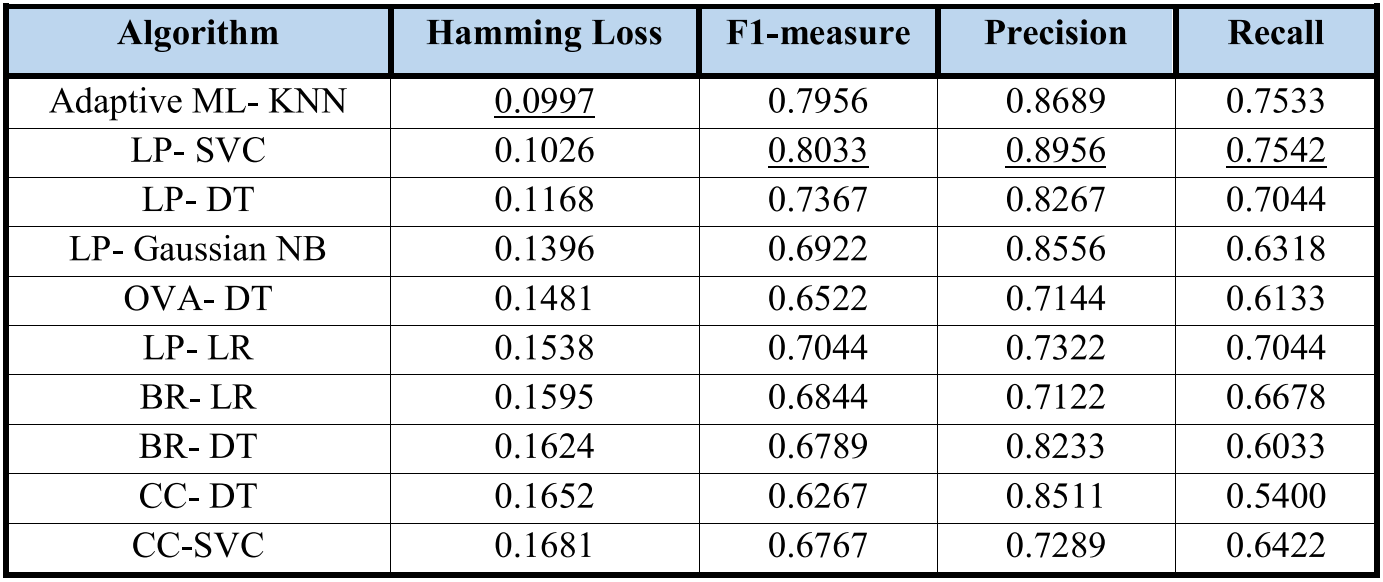
Total Rows: 100,000; Testing Rows: 20,000

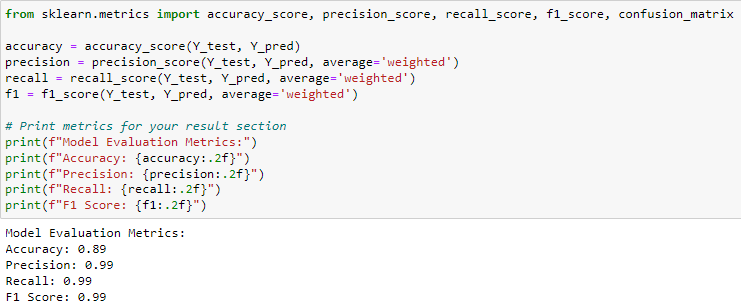
 

Metrics of Existing solutions:

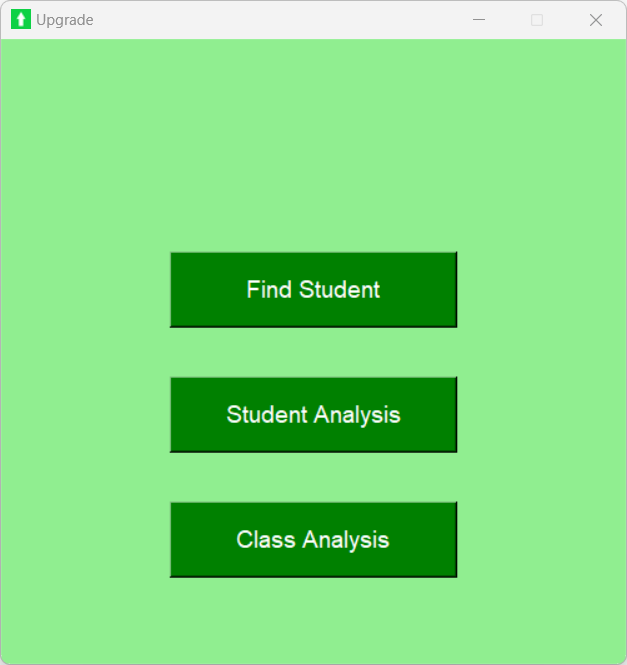
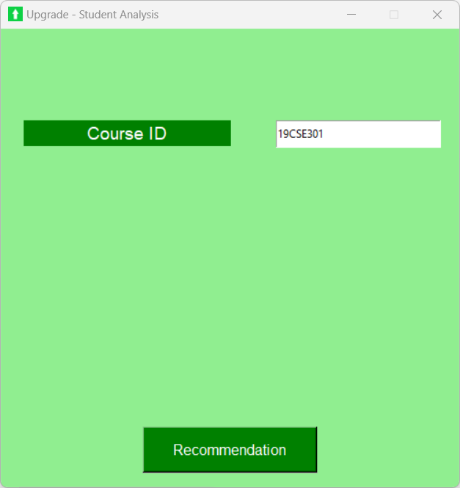
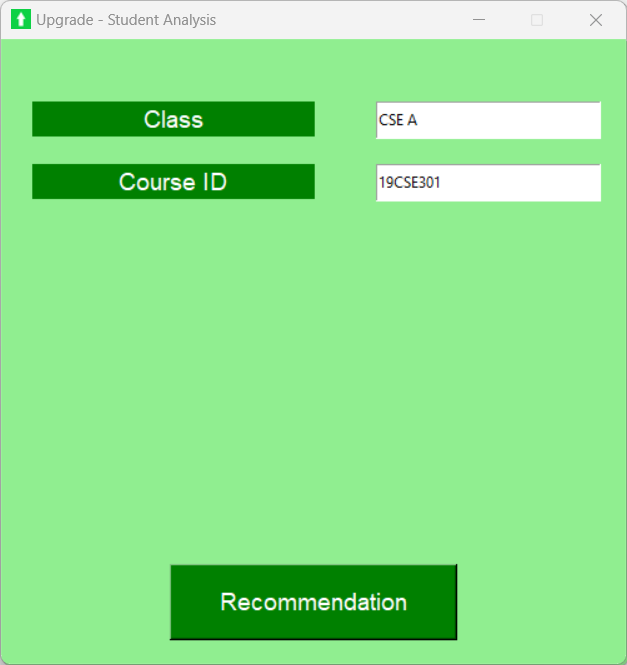


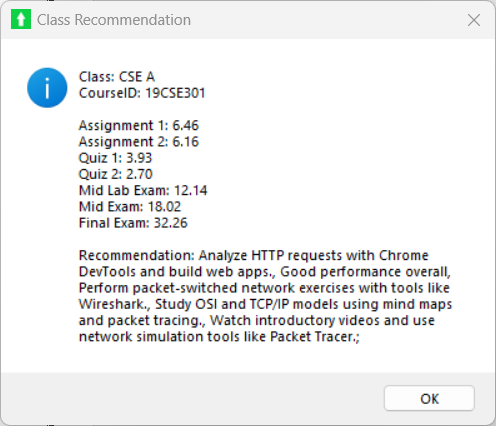
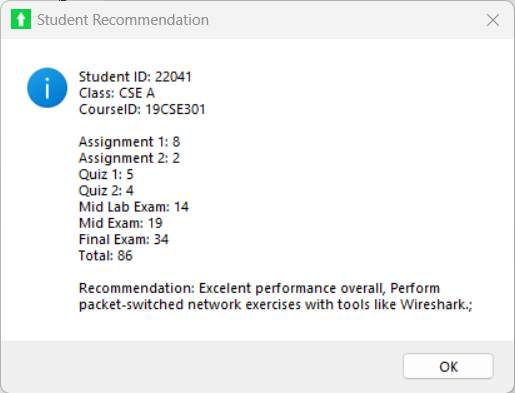
Metrics of Proposed solution:

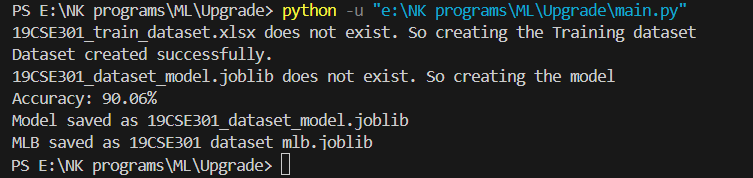


Thus, with the above results the proposed solution performs better compared to the existing solution in the terms of accuracy and generalizability.

**Sample Output:**

****  ****

** **



**Notebook File:**

<https://github.com/Nehal-Khan-29/Upgrade/blob/main/Justification%20Code%20New.ipynb>

**Conclusion:**

Thus, the proposed system is much accurate in recommending strategies which will help the students to improve their learning experience. The recommender system is more generalized, making it applicable to different courses.