University of Dundee

**Mentorship Management and Monitoring system (MMMS)**

**Submitted by: Submitted to:**

Sachin Kumar Prof. Laud Ochei

Table of Contents

1. Introduction

2. Overview of Mentorship Management and Monitoring system

3. Literature Review

4. Creation of Dataset for Mentorship Management and Monitoring system

5. Creation of Database for Mentorship Management and Monitoring system

6. Data Analysis

6.1 Description of machine learning technique for detecting conflicts in the Mentorship Management and Monitoring system (e.g., naive bayes, random forest)

6.2 Analysis of the dataset

7. Data Visualization of the result

8. Conclusion

References

Appendix

1.Introduction

Mentorship plays a vital role in supporting students academically, emotionally, and professionally. This is especially true for students who simultaneously juggle responsibilities at **universities and in companies** — having two mentors: one academic and one industrial. However, traditional mentorship management methods such as emails, informal meetings, or scattered logbooks are **inefficient, non-scalable, and lack actionable insights**.

To solve this, the **Mentor-Mentee Monitoring System (MMMS)** is designed as a smart, data-driven platform that streamlines the mentorship experience in **dual-mentorship environments**. It enables seamless coordination between students, university mentors, and company mentors by offering structured **student progress tracking**, centralized **communication**, and early **conflict detection** between academic and industrial obligations.

The core goals of MMMS include:

* Facilitating effective **mentor-mentee management** for students with dual responsibilities
* Systematically monitoring **academic performance** (attendance, assignments, etc.) alongside work commitments
* Gathering **feedback** from both academic and industry mentors
* Detecting potential conflicts (e.g., class attendance impacted by work schedules)
* Using **predictive analytics** to flag early signs of disengagement or mentorship breakdown based on data such as:
  + Communication frequency
  + Sentiment in feedback
  + Missed milestones or low academic activity

MMMS is more than just a digital tracker. It’s a **smart coordination tool** that empowers institutions and companies to provide **cohesive, proactive, and personalized mentorship support**. By making mentorship data **centralized, interpretable, and actionable**, it allows mentors and administrators to replace guesswork with **data-driven insights** — enabling more responsive decisions and stronger mentorship outcomes for students balancing both academia and industry.

This project also demonstrates how a **realistic synthetic dataset** combined with **machine learning** can be used to simulate real-world scenarios and test scalable solutions for mentorship challenges in modern educational systems.

2.Overview of MMMS

The **Mentorship Management and Monitoring System (MMMS)** is a comprehensive **platform** designed to transform how mentorship programs are managed within educational institutions — especially for students balancing **academic study and professional work**. These students often have **two mentors**: one from their **university** and another from their **company**, which introduces unique challenges in coordination, engagement, and progress tracking.

MMMS provides a centralized, data-driven, and intelligent solution that supports **continuous mentorship**, **dual-role engagement**, and **early conflict detection**, ultimately enhancing the overall mentorship experience for all stakeholders.

### **Key Features**

#### **Mentor-Mentee**

* Supports **dual-mentor assignments** (academic + industry).

#### **Progress Tracking**

* Each mentor-mentee pair (academic and industry) gets a shared digital space to:
  + Log meetings
  + Track milestone completions
* Enables **transparent accountability** and helps mentors monitor progress over time.
* The system will track student progress using quantifiable indicators such as attendance, assignment completion, and mentor engagement logs. These will serve as inputs for future ML-based risk predictions.

#### **Feedback Mechanism**

* Collected feedback feeds into **analytics models** to highlight.

#### **Conflict & Disengagement Detection**

* Employs **machine learning models** (e.g., Naive Bayes, Random Forest) to detect:
  + Missed meetings or unlogged sessions
  + Negative feedback sentiment
  + Scheduling conflicts between academic and work commitments
  + Reduced academic activity or low mentor engagement

#### **Administrative Dashboard**

* Program coordinators and academic administrators can:
  + Monitor system usage and engagement metrics
  + Identify at-risk students or ineffective mentor matches
  + Intervene early based on real-time data insights

### **Key Stakeholders**

* **Students** – Seek personalized academic and professional guidance while balancing university and work life.
* **Faculty Mentors** – Provide academic support, research supervision, and study tracking.
* **Industry Mentors** – Offer real-world exposure, soft skill development, and career coaching.
* **Program Administrators** – Oversee mentorship program health, engagement trends, and outcomes.

3.Literature Review

**Why Dual Mentorship Needs a Smarter Approach**

As more students balance **university education with industry roles**, they interact with two mentors—one academic and one professional. While this dual-mentorship model enhances real-world readiness, it also introduces challenges like **scheduling conflicts**, **misaligned expectations**, and **difficulty in tracking overall progress**.

Traditional mentorship practices—emails, spreadsheets, informal logs—are **fragmented, hard to scale, and reactive**. Without a centralized system, institutions often miss early signs of disengagement or conflict.

### **Gaps in Current Mentorship Solutions**

**Commercial platforms** (e.g., Chronus, Mentorloop) offer mentorship matching and communication tools but are geared toward corporate use. They lack:

* Support for **dual mentor structures**
* Academic data integration (e.g., attendance, assignment tracking)
* Predictive capabilities to identify issues early

**Manual systems** in academia are equally limited—inefficient, error-prone, and incapable of scaling. As noted by Crisp & Cruz (2009) and Eby et al. (2008), mentorship programs suffer when they lack structure, alignment, and continuous evaluation.

### **How AI/ML Can Transform Mentorship**

AI and machine learning enable **early detection and proactive support** in mentorship programs by:

* Analyzing **communication frequency** and **feedback sentiment**
* Predicting potential conflict using models like **Naive Bayes** and **Random Forest**
* Scoring engagement through task completion, surveys, and academic data

NLP tools can detect emotional tone in written feedback, while predictive models flag at-risk mentor-mentee pairs. Studies by Guo et al. (2019) and Sezer & Gurdal (2020) support these methods as effective in improving educational outcomes.

### **MMMS: A Data-Driven, Dual-Mentor Solution**

The **Mentorship Management and Monitoring System (MMMS)** fills a critical gap by offering:

* Unified tracking across both academic and industry mentorship
* Early conflict detection and intervention
* Personalized, data-informed support for students

By combining structured mentorship processes with AI analytics, MMMS supports a scalable, proactive, and student-centered approach to modern mentorship.

4. Creation of Synthetic Datasets

To effectively simulate the dual-mentorship environment, a collection of interconnected datasets was generated. This modular approach allows for a realistic representation of how different data points (e.g., academic performance, industry logs, meeting schedules) would be captured in a real-world system.

The following individual datasets were created as .csv files:

* **students.csv**: Contains profiles for 200 students, including their student\_id, name, email, course, and assigned academic\_mentor\_id and industry\_mentor\_id.
* **academic\_mentors.csv & industry\_mentors.csv**: These files contain profile information for 20 academic and 20 industry mentors, respectively, including their mentor\_id, name, and professional details (university or employer).
* **assignment.csv**: Tracks assignment submissions for each student, detailing the total assignments for their course, the number submitted, and a derived assignment\_grade (e.g., 'A1', 'B1', 'Fail').
* **academic\_progress.csv**: A log of each student's academic performance, including attendance percentage, the grade from the assignment.csv file, and the date of the last update.
* **industry\_progress.csv**: A log of each student's industry engagement, tracking their attendance and the number of projects they are involved in.
* **meetings.csv**: Records scheduled meetings between students and one of their mentors, capturing the meeting\_id, student\_id, mentor\_id, and date.
* **request.csv**: Logs special requests made by students, such as extensions or leave, detailing the reason and category (e.g., attendance, assignment).

#### **Master Dataset (master\_mmms.csv)**

The individual files were then merged into a single, denormalized **master dataset**. This file, master\_mmms.csv, consolidates all information for each student into a single row, providing a comprehensive view. This master file is the primary input for the data analysis and machine learning phases. It includes detailed information from student profiles, both mentor profiles, academic grades, industry logs, meeting details, and any special requests.

5. Database Schema Design

Based on the structure of the generated datasets, a relational database schema has been designed. A **SQL** database is proposed for its reliability and ability to enforce data integrity through relationships. The schema normalizes the data to reduce redundancy and ensure consistency.

The core tables in the schema are as follows:

* **master\_mms\_table**
  + Student\_id
  + Name
  + Email
  + Course
  + Current\_year
  + Term\_start\_date
  + Term\_end\_date
  + Academic\_mentor\_id
  + Academic\_mentor\_name
  + Academic\_mentor\_email
  + Academic\_mentor\_university
  + Industry\_mentor\_id
  + Industry\_mentor\_name
  + Industry\_mentor\_email
  + Industry\_mentor\_employer
  + Academic\_progress\_id
  + Number\_of\_courses
  + Total\_required\_attendance
  + Classes\_attended
  + Assignment\_id
  + Submission\_date
  + Submitted\_date
  + Assignment\_grades
  + Industry\_id
  + Hours\_per\_week
  + No\_of\_tasks
  + Academic\_meeting\_id
  + Academic\_meeting\_date
  + Industry\_meeting\_id
  + Industry\_meeting\_date
  + Request\_id
  + Num\_mitigation\_requests
  + Reason
  + category
* **students\_table**
  + student\_id (Primary Key)
  + name
  + email
  + course
  + current\_year
  + academic\_mentor\_id (Foreign Key)
  + industry\_mentor\_id (Foreign Key)
  + term\_start\_date
  + term\_end\_date
* **industry\_mentor\_table**
  + industry\_mentor\_id (Primary Key)
  + name
  + email
  + employer
* **academic\_mentor\_able**
  + academic\_mentor\_id (Primary Key)
  + name
  + email
  + university
* **assignments\_table**
  + assignment\_id (Primary Key)
  + student\_id (Foreign Key)
  + submission\_date
  + submitted\_date
  + assignment\_grades
* **academic\_progress\_table**
  + academic\_progress\_id (Primary Key)
  + student\_id (Foreign Key)
  + number\_of\_courses
  + attendance\_percentage
  + assignment
  + total\_required\_attendance
  + class\_attended
* **industry\_progress\_table**
  + industry\_id (Primary Key)
  + student\_id (Foreign Key)
  + hours\_per\_week
  + no\_of\_tasks
* **meetings\_table**
  + serial\_no((primary\_key)
  + student\_id(Foreign Key)
  + academic\_meeting\_id
  + academic\_meeting\_date
  + industry\_meeting\_id
  + industry\_meeting\_date
* **request \_table**
  + request\_id (Primary Key)
  + student\_id (Foreign Key)
  + reason
  + Category

6. Data Analysis

### **6.1 At-Risk Students Detection**

The "At-Risk" model aims to identify students likely to face challenges due to low attendance, high workload, or delayed submissions. Early identification allows mentors and administrators to intervene and support these students proactively.

**Features Used:**

* number\_of\_courses
* classes\_attended
* assignment\_grades
* no\_of\_tasks

**Target Variable:**

* Conflict\_Status (1 = at risk, 0 = not at risk)

**Model:** Random Forest Classifier with 100 estimators and balanced class weights was chosen for its robustness to imbalanced datasets and ability to capture non-linear relationships.

**Results:**

* The classification report indicates strong precision and recall for identifying at-risk students.
* The confusion matrix confirms the model effectively differentiates between at-risk and non-risk students.
* **Feature Importance:** hours\_per\_week and no\_of\_tasks are the most predictive factors, highlighting workload as a key contributor to risk.

The model effectively flags students at risk, allowing timely mentorship intervention.

**6.2 Pass/Fail Prediction**

The pass/fail model predicts whether a student will successfully pass their academic requirements, considering attendance, grades, and engagement metrics.

**Features:**

* number\_of\_courses
* classes\_attended
* hours\_per\_week
* no\_of\_tasks
* num\_mitigation\_requests

**Target Variable:**

* pass\_fail (1 = Pass, 0 = Fail)

**Model:** Random Forest Classifier, with 100 estimators and balanced class weights.

**Results:**

* The classification report demonstrates strong predictive accuracy.
* **Feature Importance:** Attendance percentage and weekly hours are key indicators of student performance.

The model provides actionable insights into which students may need additional support to ensure academic success.

### **6.3 Grade Prediction**

Linear Regression was used to predict a student’s final numeric score, which was then converted into letter grades.

**Features:**

* number\_of\_courses
* classes\_attended
* assignment\_score
* attendance\_ratio
* mitigation\_impact

**Model Performance:**

* **Mean Squared Error (MSE):** 6.51
* **Mean Absolute Error (MAE):** 2.10
* **R² Score:** 0.97

Predicted grades closely follow actual performance, validating the model’s ability to estimate student outcomes based on multiple academic and engagement metrics.

### **6.4 Engagement & Conflict Features**

Student engagement and potential conflicts were assessed using mentorship meeting attendance, communication gaps, and conflict-related keywords.

**Features:**

* communication\_gap
* academic\_mentor\_id
* industry\_mentor\_id
* academic\_mentor\_university
* industry\_mentor\_employer

**Target Variable:**

* Conflict (1 = conflict, 0 = no conflict)

**Model:** Random Forest Classifier

**Results:**

* The classification report confirms the model effectively detects conflicts arising from cancelled meetings or miscommunication.
* **Feature Importance:** Communication gaps and mentor identifiers play a significant role in predicting conflicts.

These insights help administrators address engagement issues and improve mentorship alignment between academic and industry mentors.

### **6.5 Mitigation Request Patterns**

Apriori analysis was applied to student mitigation requests to identify frequent patterns and potential areas of concern.

**Methodology:**

* Transaction encoding of student requests
* Frequent itemset mining (support ≥ 0.1)
* Association rules generated (confidence ≥ 0.6)

**Results:**

* **Frequent Itemsets:** Identifies the most common reasons for mitigation requests.
* **Association Rules:** Reveals co-occurring reasons, providing insights into recurring challenges.

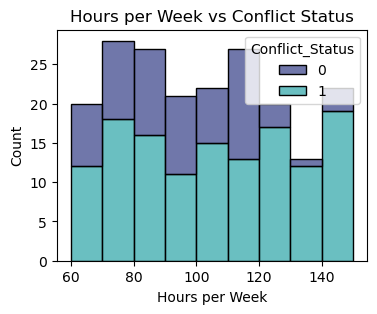
7. Data Visualization of Results

Data visualization provides a clear, interpretable way to understand student performance, engagement, conflict risks, and mentorship patterns. The visualizations below illustrate key insights derived from the Week 3 analysis.

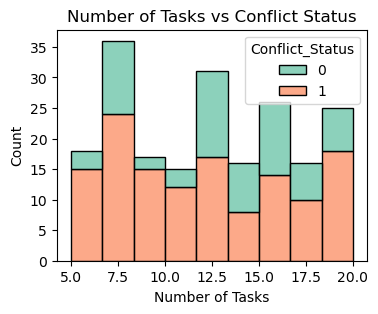
### 7.1 At-Risk Students

**Visualization:**

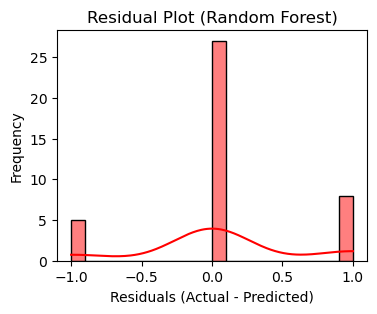
**Hours per Week vs Conflict Status** – shows students with higher weekly hours are more likely to be at risk.



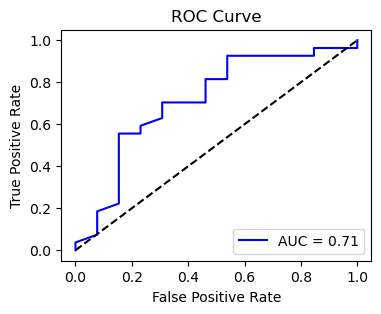
**Number of Tasks vs Conflict Status** – indicates a higher workload correlates with increased conflict probability.



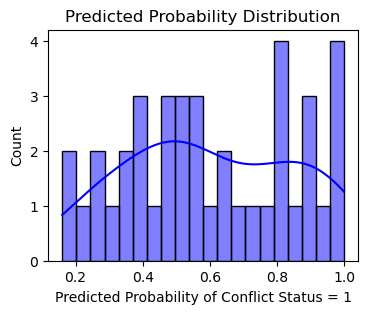
**Residual Plot:** Shows the differences between actual and predicted conflict labels.



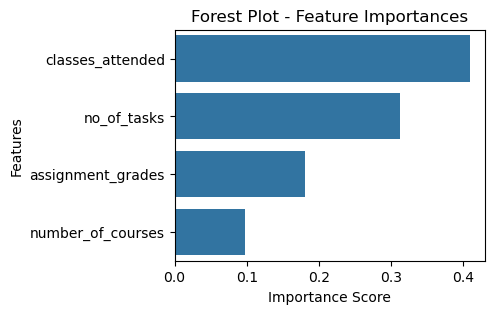
**ROC Curve:** Evaluates how well the model distinguishes between at-risk and not-at-risk students.



**Predicted Probability Distribution:** Shows predicted probabilities of a student being at risk.



**Forest plot:** Shows which factors the Random Forest model considered most important in predicting whether a student is at risk.

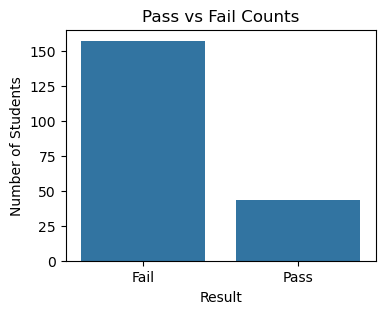


**Interpretation:** These visualizations confirm that workload management is crucial. Students with many tasks or long work hours often struggle to maintain balanced academic and industry responsibilities.

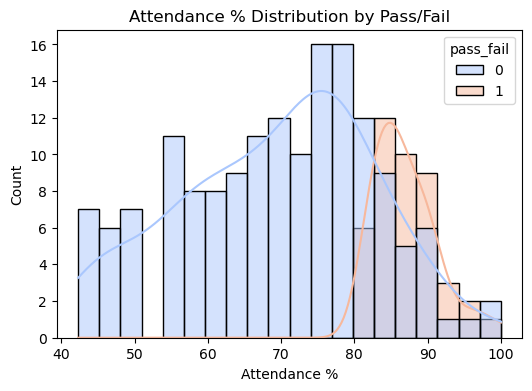
### 7.2 Pass/Fail Analysis

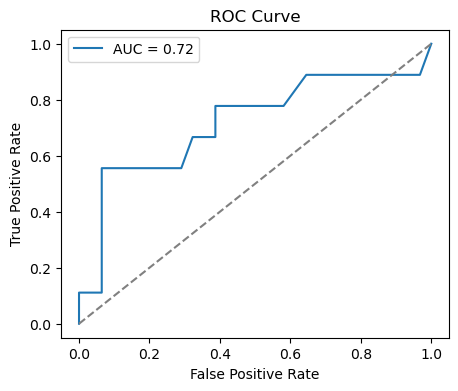
**Visualization:**

**Pass vs Fail Counts** – illustrates the overall distribution of academic success.



**Attendance % Distribution by Pass/Fail** – demonstrates that higher attendance strongly correlates with passing grades.



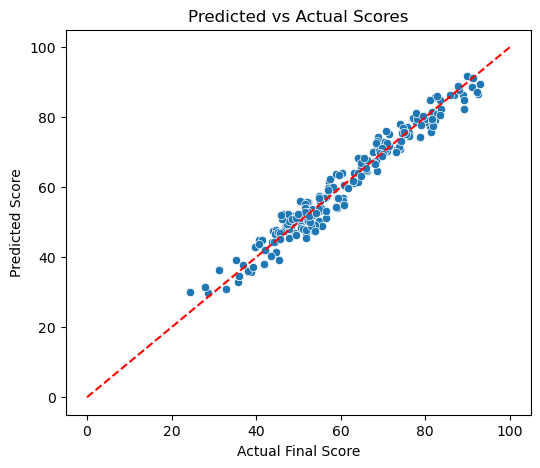
**ROC Curve** – shows that the pass/fail predictive model effectively discriminates between students who pass and fail.  


**Interpretation:** Attendance, task completion, and mitigation request management are critical indicators of student success. These metrics allow mentors to identify at-risk students proactively.

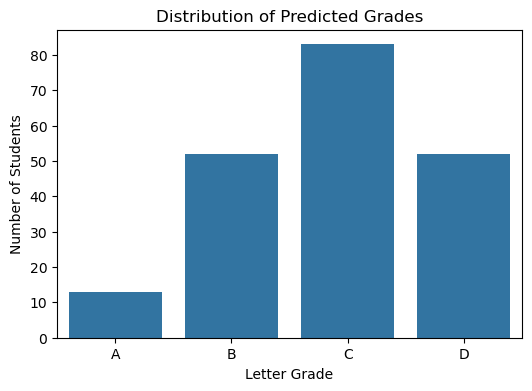
### 7.3 Grade Prediction

**Visualization:**

**Predicted vs Actual Scores Scatterplot** – confirms high alignment between predicted and actual final scores.



**Distribution of Predicted Letter Grades** – highlights expected grade distributions and identifies students who may require additional support.

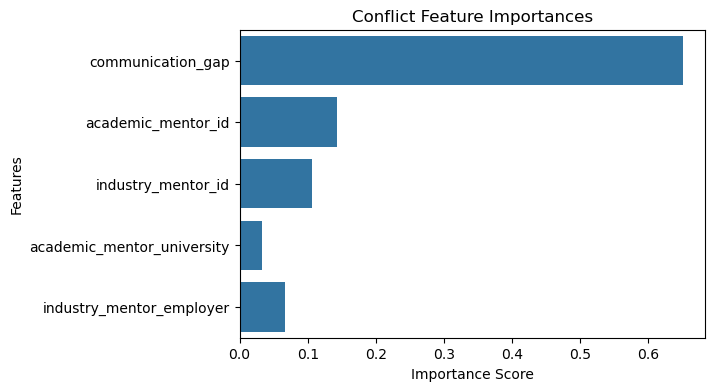


**Interpretation:** The linear regression model reliably predicts numeric scores and converts them into letter grades. This helps mentors forecast academic performance and guide interventions.

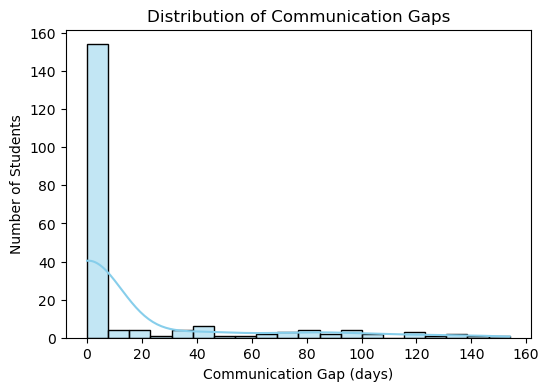
### 7.4 Engagement & Conflict Detection

**Visualization:**

**Barplot of Feature Importances** – shows which factors most influence conflict detection.



**Communication Gap Analysis** – highlights students with delayed or missing mentor interactions.

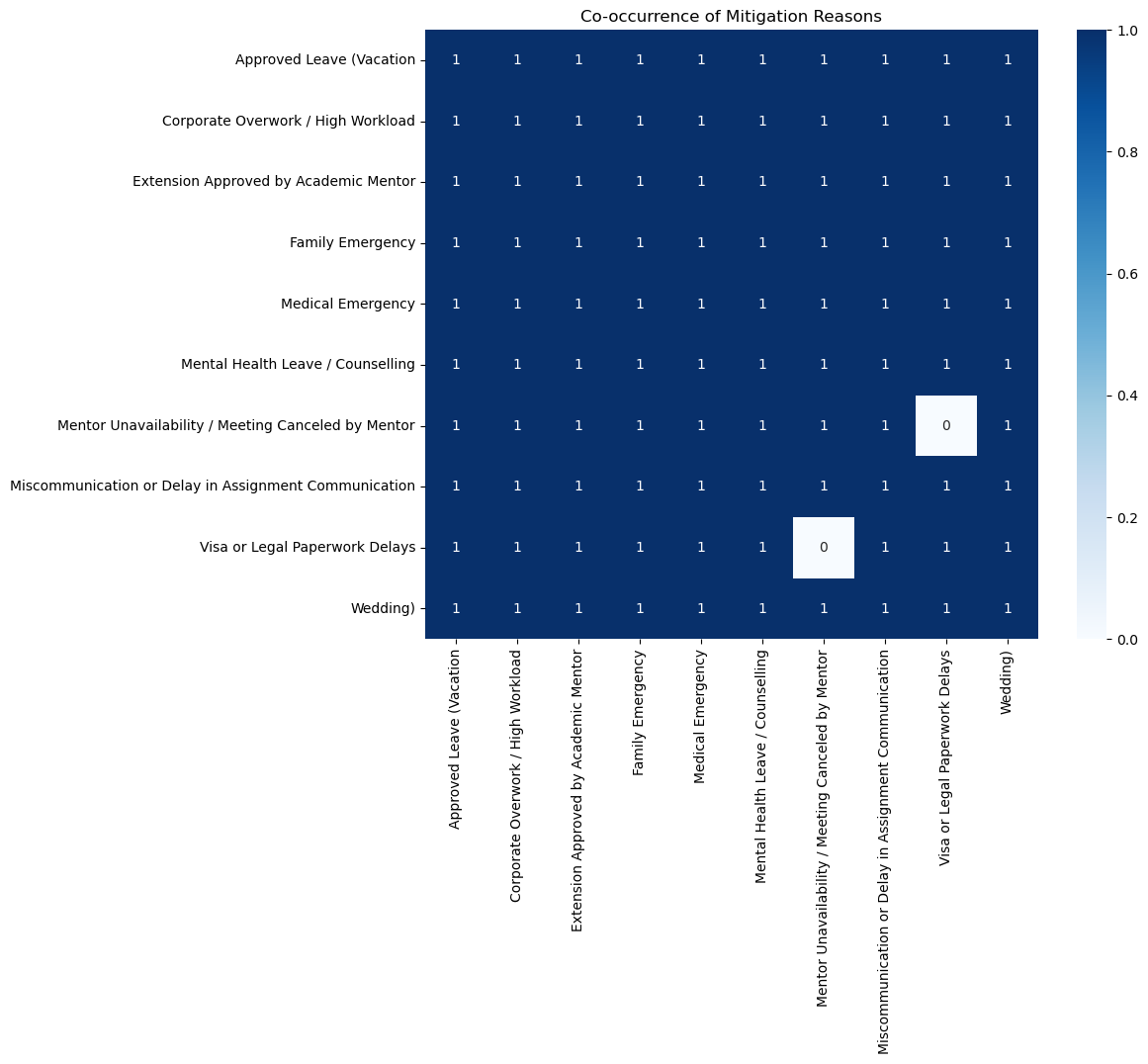


**Interpretation:** Communication gaps and canceled meetings are primary predictors of mentorship conflicts. Early detection allows administrators to intervene and align mentor-student expectations.

### 7.5 Mitigation Request Patterns

**Visualization:**

* **Co-occurrence Heatmap of Mitigation Reasons** – illustrates common combinations of mitigation requests, such as “delay” with “unavailability” or “miscommunication”.



**Interpretation:** Recurring mitigation requests indicate areas where program improvements are needed, e.g., scheduling, task allocation, or clearer communication between mentors and students.

## **8. Conclusion**

The Mentorship Management and Monitoring System (MMMS) demonstrates the value of a data-driven, dual-mentor framework for students balancing academic and professional responsibilities. Through Weeks 1–4, the project achieved the following:

1. **Comprehensive Dataset Creation:**
   * Generated realistic synthetic datasets capturing student profiles, mentor feedback, attendance, tasks, meetings, and requests.
   * Consolidated datasets into a master dataset, supporting machine learning analysis.
2. **Database Implementation:**
   * Designed a normalized relational database for robust data storage, supporting queries on students, mentors, assignments, meetings, and requests.
3. **Predictive Modeling:**
   * Random Forest models accurately identified at-risk students and predicted pass/fail outcomes.
   * Linear Regression effectively predicted final grades.
   * Conflict detection models flagged students with engagement issues or mentorship misalignments.
   * Apriori analysis revealed frequent patterns in mitigation requests, informing potential program improvements.
4. **Data Visualization:**
   * Graphical representations highlighted workload, attendance, engagement, and conflict patterns.
   * Feature importance plots emphasized key drivers of academic performance and mentorship conflicts.
5. **Insights & Recommendations:**
   * High workloads and missed meetings are major predictors of conflict and poor performance.
   * Early intervention is critical to prevent student disengagement.
   * Monitoring communication gaps and mitigation requests allows proactive mentor support.
   * Future enhancements could integrate NLP on mentor feedback, real-time alerts, and adaptive scheduling algorithms.

In conclusion, MMMS provides a scalable, intelligent platform for monitoring and enhancing dual mentorship programs. By combining structured tracking with predictive analytics, the system enables administrators and mentors to make data-informed decisions that improve student outcomes, reduce conflicts, and foster successful academic and professional development.

References

1. Crisp, G., & Cruz, I. (2009). Mentoring college students: A critical review of the literature between 1990 and 2007. *Research in Higher Education, 50*(6), 525–545.
2. Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior, 72*(2), 254–267.
3. Guo, S., et al. (2019). Applying machine learning for academic performance prediction in higher education: A systematic review. *Computers & Education, 142*, 103644.
4. Sezer, E. A., & Gurdal, M. (2020). Data-driven student performance modeling using machine learning: Predicting risk and engagement. *Education and Information Technologies, 25*(2), 1011–1030.

Appendix

* Synthetic dataset files (.csv)
* SQL database dump (.sql)
* Python scripts for dataset generation and ML models (.ipynb/.py)
* All visualizations (.png)