1. Overview

This document presents the interpretation of student clusters identified using **K-Means clustering** as part of **Activity 3: Machine Learning – Student Segmentation & Risk Detection**.

Clustering was performed on scored features derived in Activity 2:

- Academic Performance Score (APS)
- Wellness & Wellbeing Score (WWS)
- Productivity & Time Management Score (PTMS)
- Career Readiness Score (CRS)
- Student Readiness Index (SRI)

The objective of clustering is to uncover **hidden student segments** and identify **risk and opportunity groups** that support targeted mentoring interventions under the HEPro Dedicated Mentoring System.

Based on the project objective, students were segmented into **three meaningful clusters**:

- At-Risk Students
- High Performers
- Career-Confused Students

---

2. Cluster-wise Score Summary

Table 1 shows the average scores for each cluster across the key dimensions used for segmentation.

Table 1: Average Scores by Cluster

Cluster Label	APS	WWS	PTMS	CRS	SRI
At-Risk Students	68.43	53.09	39.69	52.42	54.85
Career-Confused Students	73.43	71.44	57.41	40.99	69.12
High Performers	73.35	52.63	70.21	44.40	60.32

Interpretation

- **At-Risk Students** show comparatively low wellbeing and productivity, resulting in the lowest overall readiness.
- **Career-Confused Students** demonstrate strong academic ability but significantly lower career readiness, which limits their overall preparedness.
- **High Performers** exhibit strong academic and productivity scores, leading to higher readiness levels.

3. Cluster Interpretation

Cluster 1: At-Risk Students

Key Characteristics

- Low APS, WWS, PTMS, CRS, and SRI
- High stress and low productivity
- Weak academic consistency

Interpretation
Students in this cluster experience academic challenges combined with poor wellbeing and time management. These factors collectively reduce their readiness score. This group represents the highest-risk segment and requires immediate mentoring intervention.

Cluster 2: High Performers

Key Characteristics

- High APS and PTMS
- Strong consistency and engagement
- Higher overall SRI

Interpretation
These students demonstrate strong academic performance, effective productivity, and relatively stable wellbeing. They require minimal intervention and can be supported through advanced mentoring, leadership development, or peer mentoring opportunities.

Cluster 3: Career-Confused Students

Key Characteristics

- High APS
- Moderate to low CRS
- Medium SRI
- Good academic ability but unclear career direction

Interpretation
Students in this cluster perform well academically but lack clarity regarding career goals or skill alignment. Without targeted career guidance, these high-potential students may face long-term direction issues.

---

4. Comparison with Rule-Based SRI Categories

To validate machine learning insights, the identified clusters were compared with rule-based SRI categories.

Table 2: ML Cluster vs Rule-Based SRI Category

Cluster Label	Blue	Green	Red	Yellow
At-Risk Students	4	0	1	14
Career-Confused Students	8	0	0	7
High Performers	14	2	0	0

Interpretation

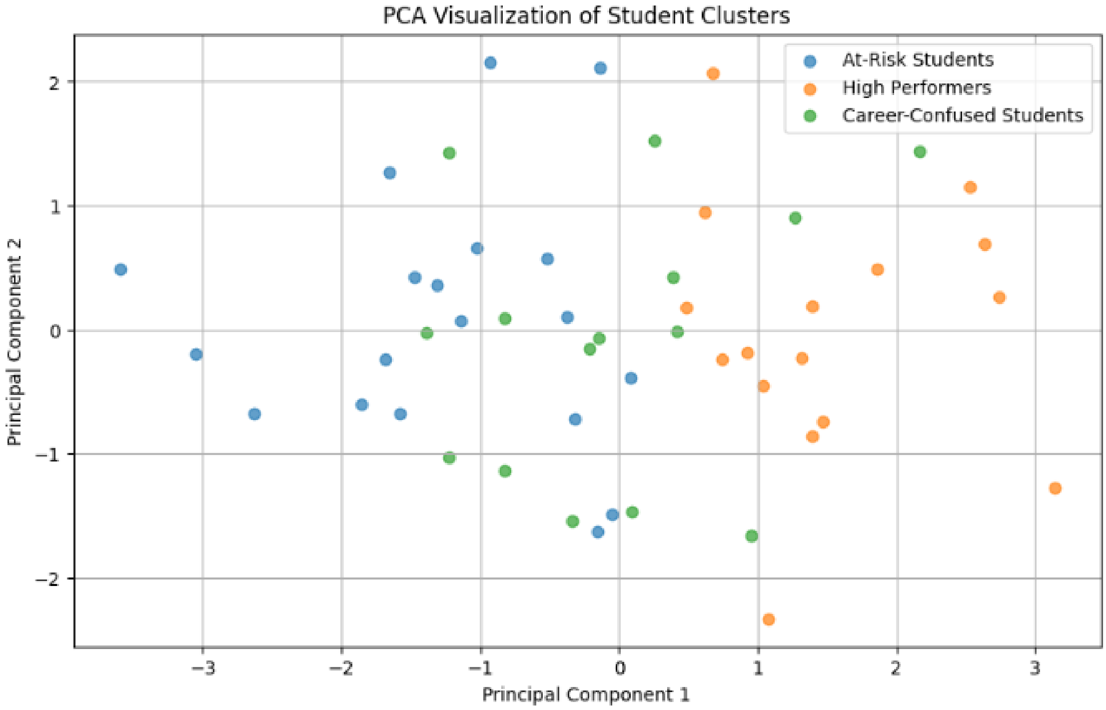
- Most **At-Risk Students** fall under Yellow and Red categories, confirming high intervention needs.
- **Career-Confused Students** are largely classified as Blue and Yellow, indicating moderate readiness with specific career gaps.
- **High Performers** are primarily classified as Blue and Green, showing strong alignment between ML-based and rule-based evaluation.

This comparison demonstrates that ML-based clustering complements the rule-based SRI by revealing nuanced patterns.

---

5. Cluster Visualization

A **PCA-based visualization** was used to project high-dimensional student scores into two dimensions for interpretability.



The visualization shows clear separation between:

- At-risk students
- High performers
- Career-confused students

This confirms that the clusters identified by K-Means are meaningful and well-separated.

---

6. Conclusion

The clustering results reveal distinct student groups with different mentoring needs. Machine learning–based segmentation enhances the rule-based scoring system by uncovering patterns that are not immediately visible through SRI alone. This combined approach enables more personalized and effective mentoring interventions under the HEPro framework.