# Sentence Encoder-Based Clustering Method for Modeling Students' Learning Programming Behavior

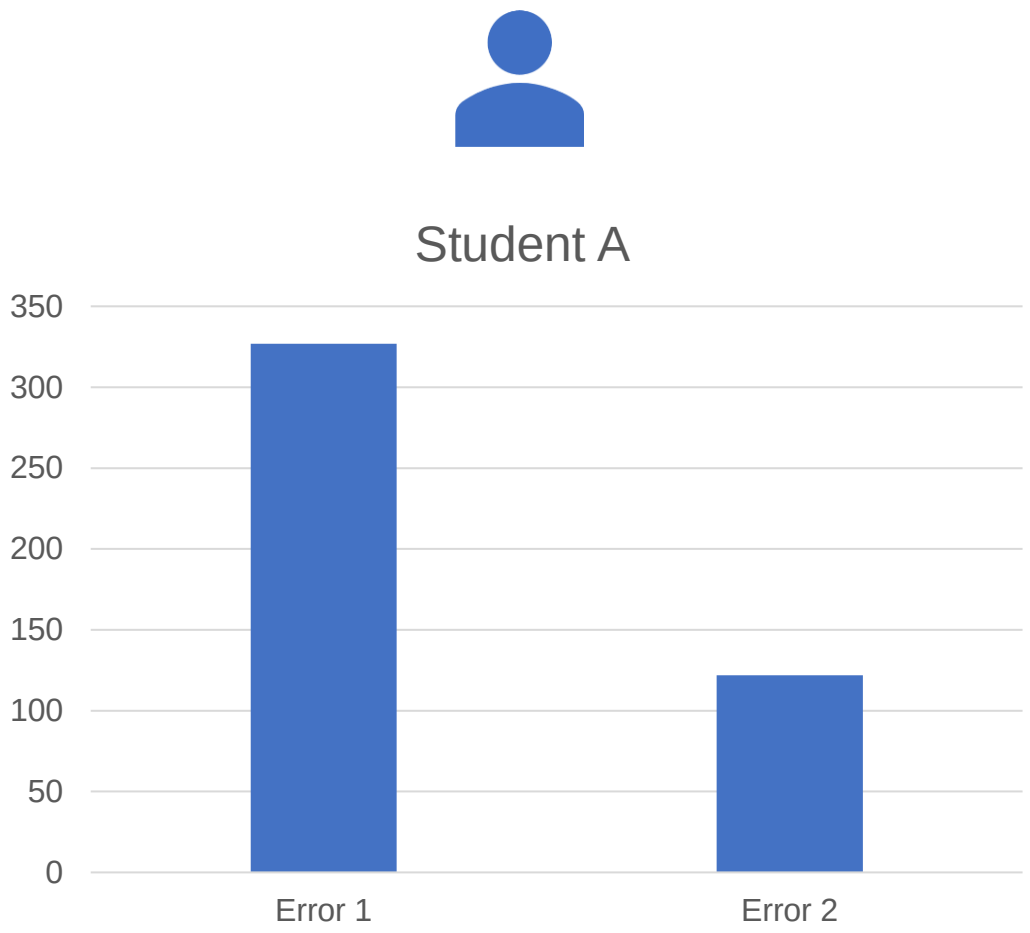
**Mubina Kamberovic**[1,2], Amina Mevic[1], Senka Krivic[1]

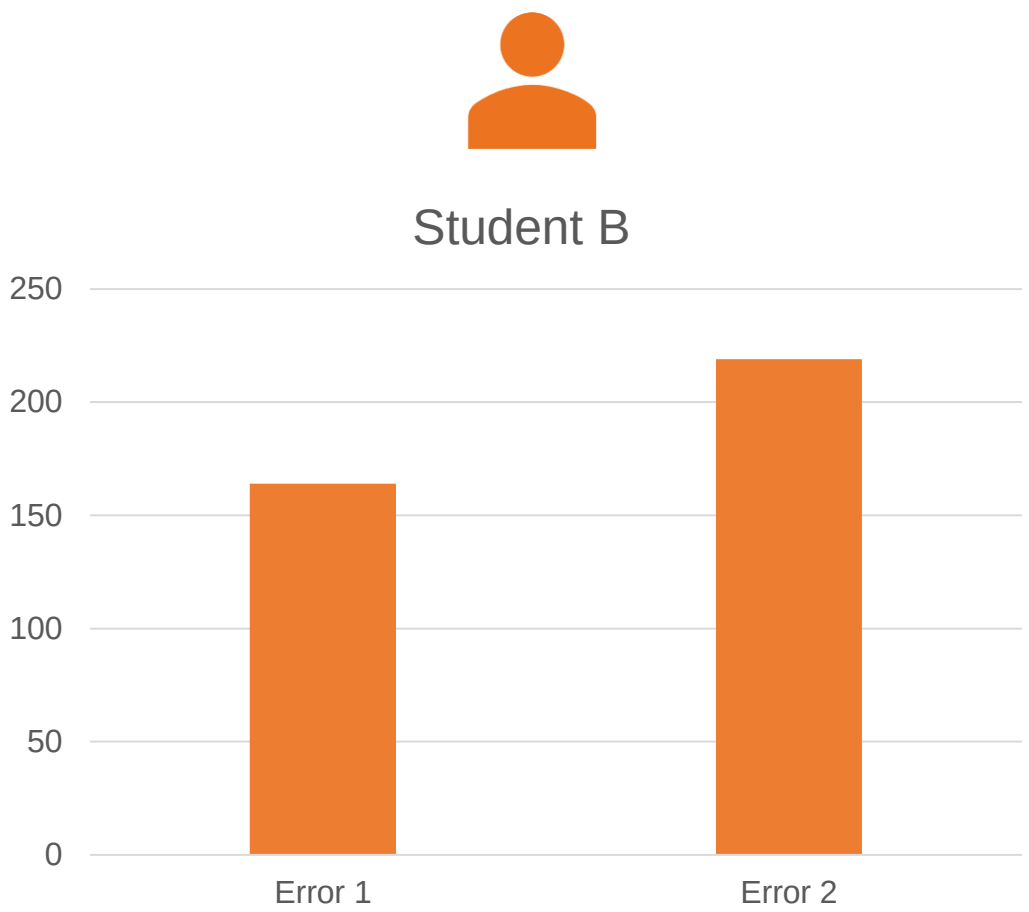
[1] Faculty of Electrical Engineering, University of Sarajevo

[2] Olivebh, Sarajevo

33rd ACM International Conference on User Modeling, Adaptation and Personalization

# **3 out of 10** students‘ compilations end with failure



# Research Questions

**RQ1.** Could compiler errors help us predict performance and future errors?

**RQ2.** Could prior clustering of students enhance the predictions?

**RQ 3.** When can meaningful student clusters be identified early enough to improve predictions?

# State of the Art

Both performance and errors prediction

In a programming environment

**Performance Prediction**

- Based on compiler errors (Rodrigo et al. 2009)
- K-Means clustering prior to training distinct classifiers improved predictions (Al-Tameemi et al. 2024)

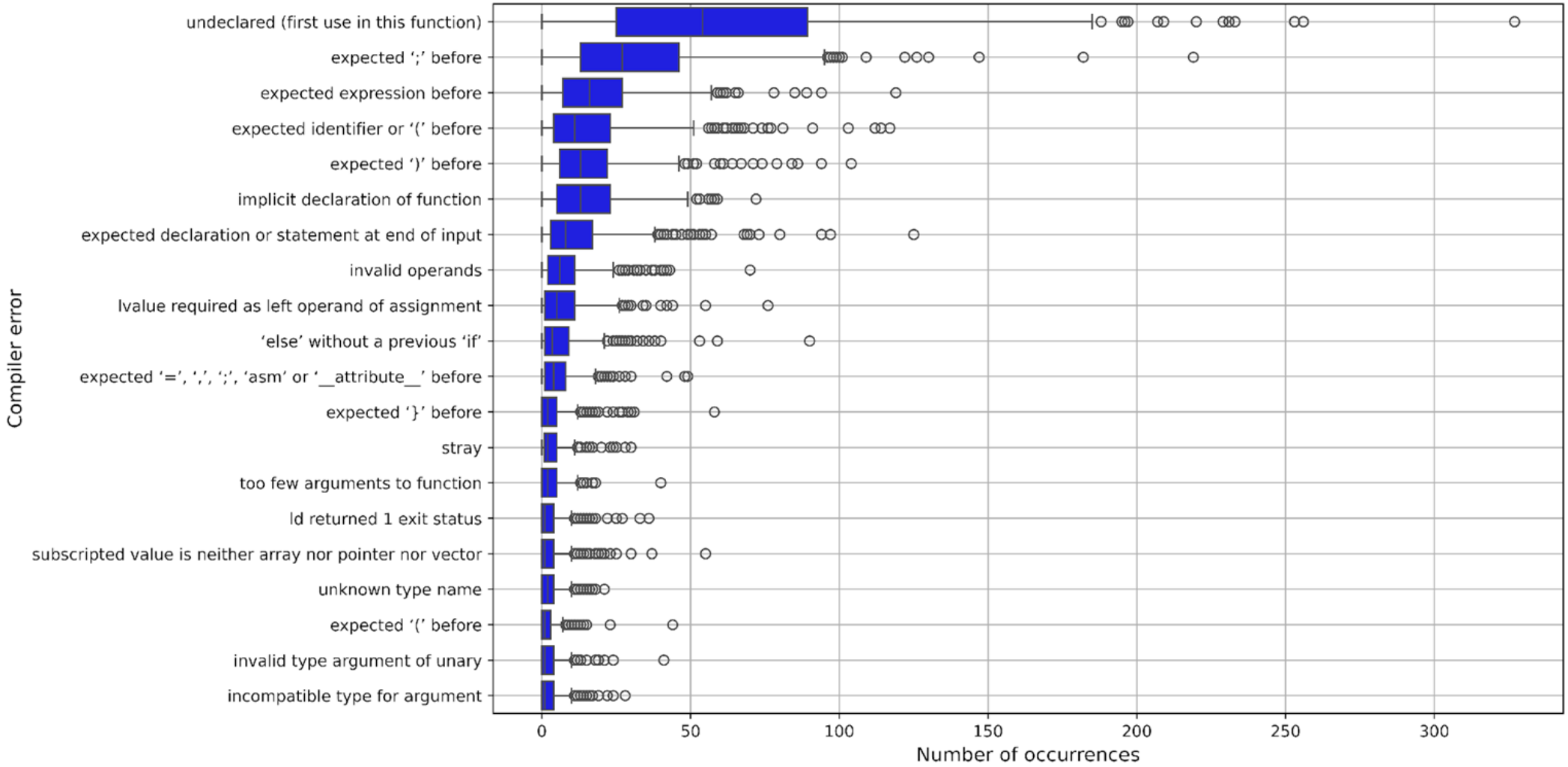
Regression

**Clustering**

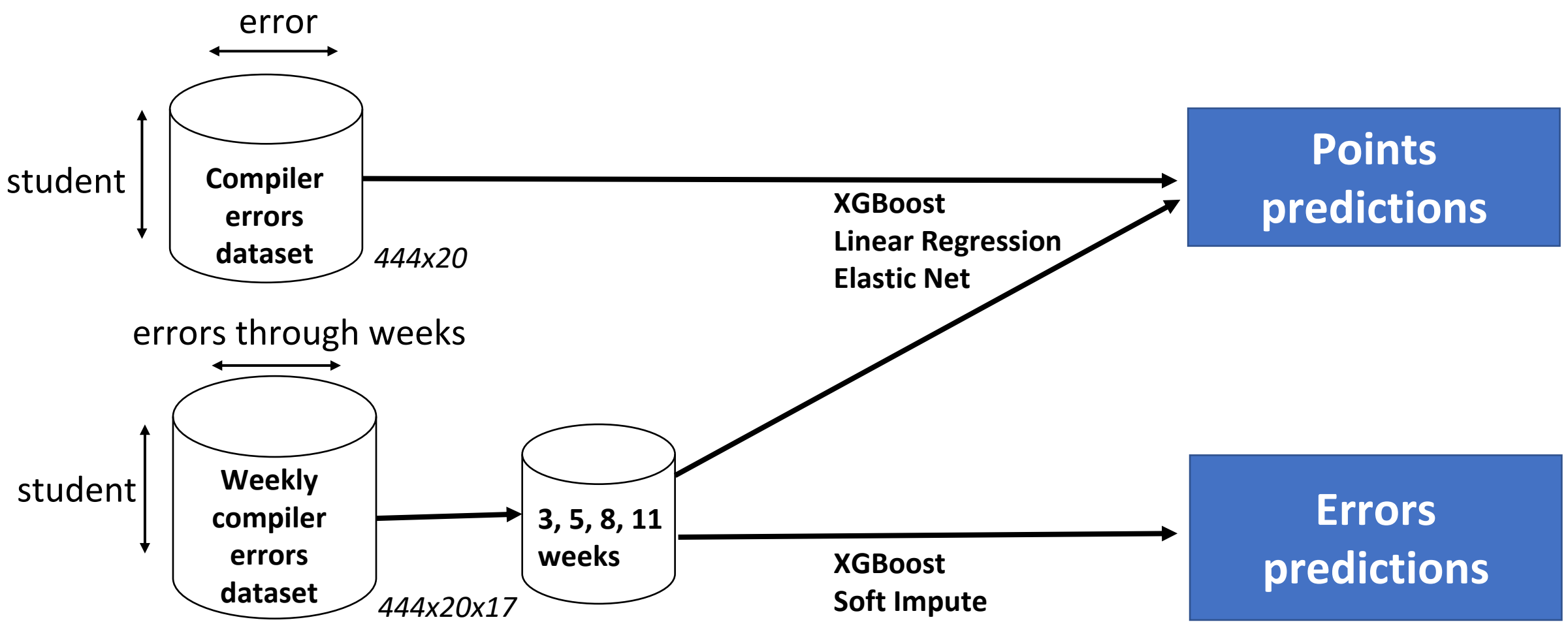
- Based on behavioral patterns from question-based learning environments (Mirzaei et al. 2019)
- K-Means on the sentence-encoded dataset in banking and e-commerce domains (Tissera et al. 2024)

# Dataset

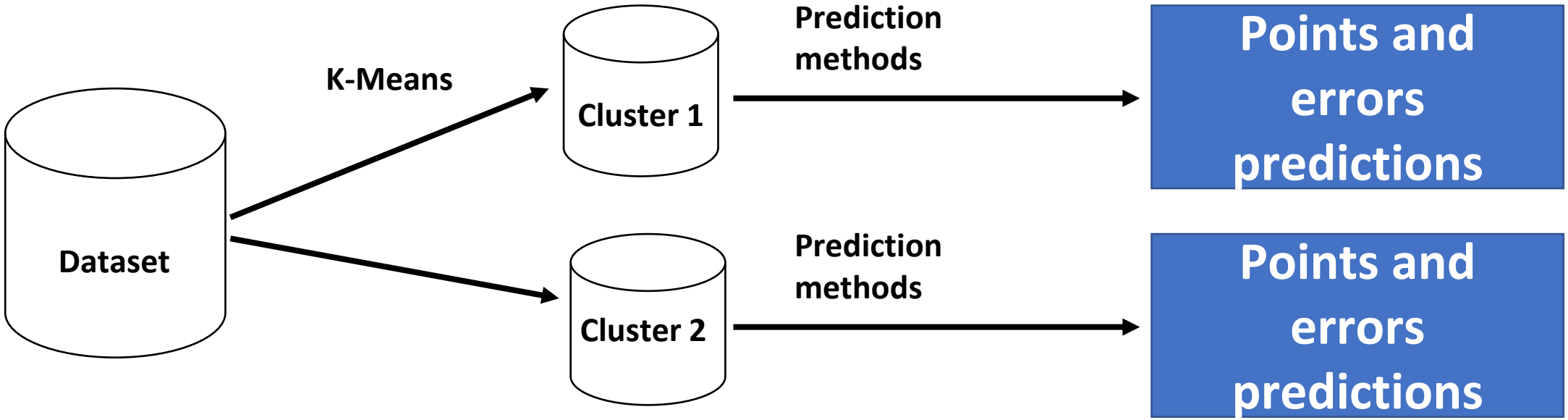
Logged data for 444 students from the C9 programming environment



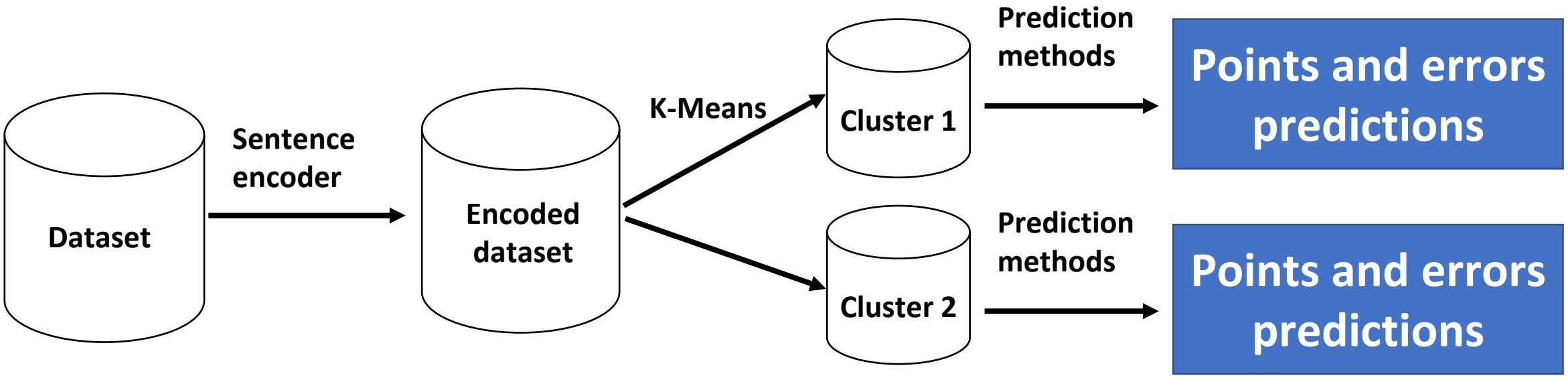
# Baseline Predictions



# Clustering before Prediction

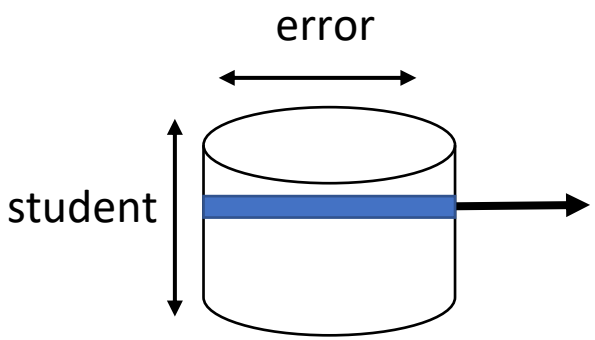


# Clustering on Sentence-Encoded Dataset



# Sentence Encoding

***Raw dataset***




***Paragraph Generation***

*For the summed dataset:*
Student made the ‘error 0’ 78 times, the ‘error 1’ 37 times, the ‘error 2’ 17 times, the ‘error 3’ 29 times, the ‘error 4’ 19 times, the ‘error 5’ 19 times...

*For the weekly dataset:*
Student made ‘error 0’ 5 times, ‘error 3’ 5 times, ‘error 5’ 2 times, ‘error 6’ 1 time, ‘error 7’ 3 times, ‘error 13’ 5 times in this week. In previous weeks, student made 201 errors.

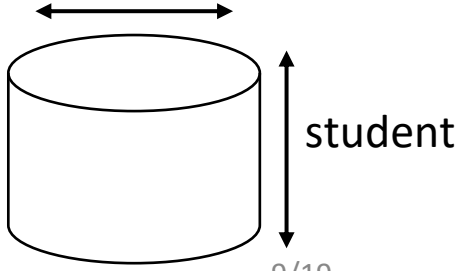
***paraphrase-multilingual-mpnet-base-v2***
(Reimers and Gurevych, 2019)

768-dimensional vector

[-1.06682412e-01, -2.10228980e-01,...]

encoded paragraph

***Encoded dataset***



# RQ1. Could compiler errors help us predict performance and future errors?

# Definitely!

*Best points prediction*

- **Sum of all errors**
- RMSE:**0.191**, MAE:**0.16**
- Model: **XGBoost**

*Best future errors prediction*

- **After 5 weeks**
- RMSE:**0.098**, MAE:**0.055**
- Model: **XGBoost**

# Errors Importance for **Points** Prediction

undeclared (first use in this function)

expected ;

implicit declaration of function

# Errors Importance for **Errors** Prediction

undeclared (first use in this function)

expected ;

# RQ2. Could prior clustering of students enhance the predictions?

# Points prediction accuracy based on **weekly** errors

| Clustering method | Prediction method | 3 weeks (RMSE) | 3 weeks (MAE) | 5 weeks (RMSE) | 5 weeks (MAE) | 8 weeks (RMSE) | 8 weeks (MAE) | 11 weeks (RMSE) | 11 weeks (MAE) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **Points Prediction** | | | | | |
| No clustering | XGBoost | 0.252 | 0.223 | 0.250 | 0.211 | 0.224 | 0.185 | 0.200 | 0.156 |
| | LR | 0.727 | 0.360 | >1 | >1 | >1 | >1 | >1 | >1 |
| | Elastic Net | 0.248 | 0.221 | 0.248 | 0.22 | 0.242 | 0.213 | 0.231 | 0.202 |
| K-Means | XGBoost | 0.239 | 0.201 | 0.250 | 0.223 | **0.186** | **0.163** | **0.178** | **0.145** |
| | LR | 0.281 | 0.238 | >1 | >1 | 0.661 | 0.408 | >1 | >1 |
| | Elastic Net | **0.227** | **0.197** | 0.253 | 0.228 | 0.208 | 0.188 | 0.214 | 0.187 |
| SE K-Means | XGBoost | 0.240 | 0.204 | **0.202** | **0.181** | 0.254 | 0.206 | 0.210 | 0.178 |
| | LR | >1 | >1 | 0.228 | 0.199 | >1 | >1 | >1 | >1 |
| | Elastic Net | 0.239 | 0.202 | 0.210 | 0.189 | 0.247 | 0.219 | 0.217 | 0.195 |

# Errors prediction accuracy based on **weekly** errors

| Clustering method | Prediction method | 3 weeks (RMSE) | 3 weeks (MAE) | 5 weeks (RMSE) | 5 weeks (MAE) | 8 weeks (RMSE) | 8 weeks (MAE) | 11 weeks (RMSE) | 11 weeks (MAE) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **Errors Prediction** | | | | | |
| No clustering | XGBoost | 0.100 | 0.058 | 0.098 | 0.055 | 0.100 | 0.056 | 0.099 | 0.056 |
| | Soft Impute | 0.116 | 0.044 | 0.113 | 0.045 | 0.108 | 0.049 | 0.106 | 0.051 |
| K-Means | XGBoost | 0.104 | 0.068 | 0.108 | 0.067 | 0.105 | 0.066 | 0.118 | 0.071 |
| | Soft Impute | 0.130 | 0.054 | 0.128 | 0.058 | 0.117 | 0.056 | 0.130 | 0.066 |
| | XGBoost | **0.079** | 0.047 | **0.071** | 0.041 | **0.092** | 0.051 | 0.086 | 0.050 |
| SE K-Means | Soft Impute | 0.102 | **0.037** | 0.091 | **0.034** | 0.105 | **0.044** | **0.079** | **0.035** |

RQ3. When can meaningful student clusters be identified early enough to improve predictions?

# Cluster Analysis

| Weeks | Clustering method | | Size | Average errors | Average points |
|---|---|---|---|---|---|
| 11 | SE K-Means | Cluster 1 | 72 | 67.681 | 27.453 |
| | | Cluster 2 | 372 | 267.417 | 41.665 |
| | K-Means | Cluster 1 | 334 | 155.398 | 38.462 |
| | | Cluster 2 | 110 | 476.809 | 42.088 |
| 5 | SE K-Means | Cluster 1 | 52 | 49.154 | 36.504 |
| | | Cluster 2 | 392 | 259.684 | 39.739 |
| | K-Means | Cluster 1 | 108 | 450.787 | 41.904 |
| | | Cluster 2 | 336 | 165.676 | 38.543 |
| 3 | SE K-Means | Cluster 1 | 358 | 271.117 | 39.871 |
| | | Cluster 2 | 86 | 84.791 | 37.233 |
| | K-Means | Cluster 1 | 64 | 396.953 | 35.350 |
| | | Cluster 2 | 380 | 207.755 | 40.036 |

# Conclusion

Compiler errors are good performance indicators

The most frequent errors are not necessarily the most important ones

Clustering the sentence encoded dataset enhances predictions, identifying a group of students with high and low total error count

Thank you for your attention!


Mubina Kamberovic

Contact:
mkamberovi1
@etf.unsa.ba

Amina Mevic

Contact:
amevic
@etf.unsa.ba

Senka Krivic

Contact:
senka.krivic
@etf.unsa.ba


Presentation slides will be available at: