Carnegie Mellon University
Master of Computational Data Science

Carnegie Mellon University
Language Technologies Institute

Predicting Learning Outcomes

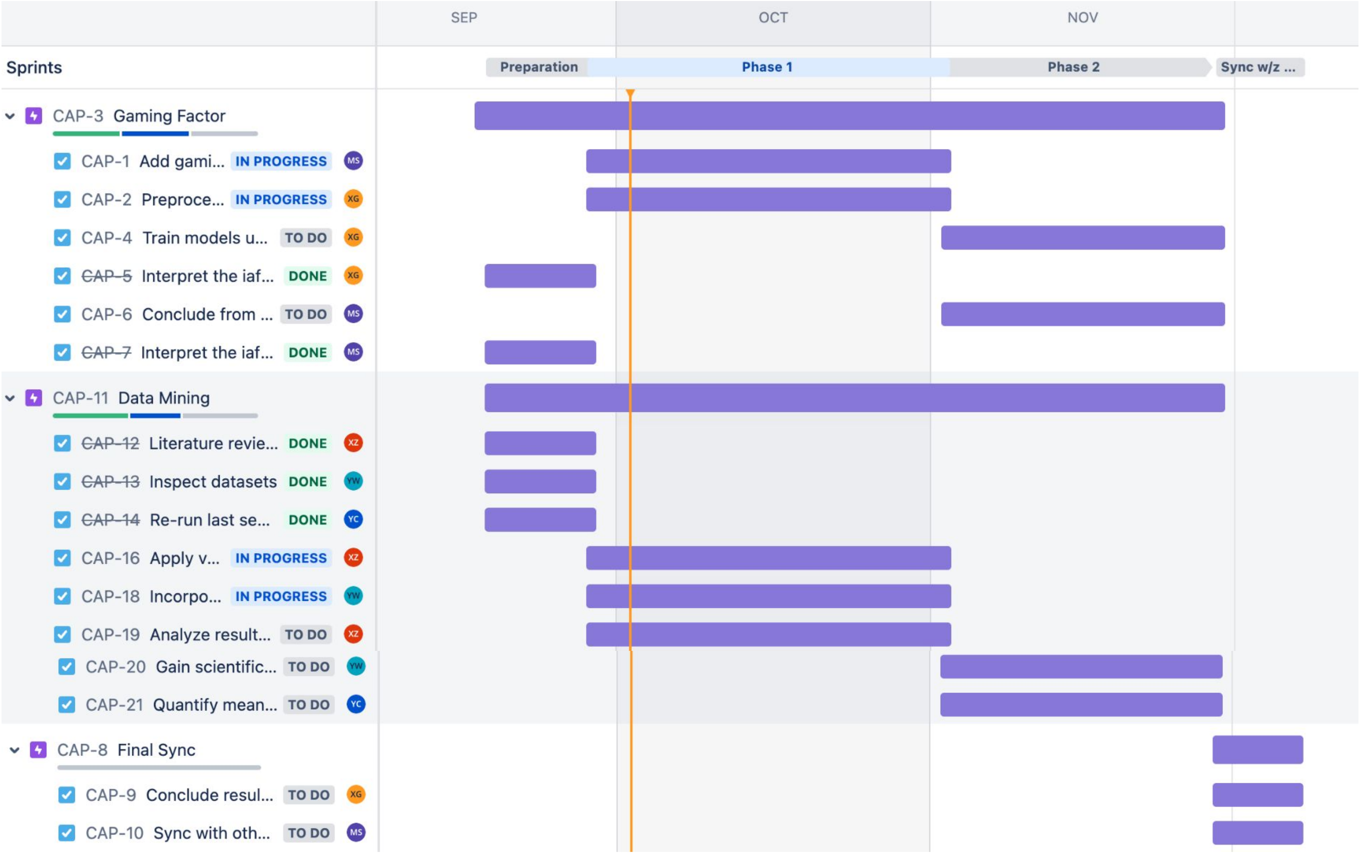
Oct 2 Standup

Mengjie Shen, Xiaoyu Zhang, Xinyu Gu, Yizhou Chen, Yuchen Wang

Advisor: Ken Koedinger

11-632 (Fall 2023)
MCDS Capstone Course

Fall Timeline



Subgroup 1 - Xinyu Gu, Mengjie Shen

Last week:

- Revisit existing model results with mentors
  - Check input data quality
  - Check model formulation and interpretation
- Complete literature review on gaming factors and decide next steps
- Finalize Fall schedule with mentors

This week:

- Add gaming factor into existing models and interpret significance - Mengjie
- Explore a new dataset (called Manuf 5165) on students' learning behaviors and process data for future analysis - Xinyu

Subgroup 2 - Interpret and improve existing models

Stage 1: Predict correctness of a problem in tutoring system

```
model_iafm = sample_data %>%
  glmer(response ~ opportunity0 + (opportunity0|KC) + (opportunity0|individual),
        data=., family=binomial(), nAGQ = 0 )
summ(model_iafm)
```

Model: generalized linear **mixed-effects** (fixed and random effects) model

Each group has a set of intercept and slope

Features used: `KC` - knowledge component, indicates type of the problem

`opportunity0` - # of times a student has answer the corresponding type of problem

`individual` - student identification

Stage 1: Predict correctness of a problem in tutoring system

```
model_iafm = sample_data %>%
  glmer(response ~ opportunity0 + (opportunity0|KC) + (opportunity0|individual),
        data=., family=binomial(), nAGQ = 0 )
summ(model_iafm)
```

`Intercepts of KC` - easiness of a knowledge component

`Intercepts of individual` - initial knowledge of a student

```
model_iafm_reverse = sample_data %>%
  glmer(response ~ opportunity_reverse + (opportunity_reverse|KC) + (opportunity_reverse|individual),
        data=., family=binomial(), nAGQ = 0 )
summ(model_iafm_reverse)
```

opportunity_reverse = max(opportunity) - opportunity

`Intercepts of individual (opportunity_reverse)` -
an estimate of the student's knowledge at the end of the learning process

Stage 2: Predict final math score

```
testScoresAll %>%
    lm(FinalMath ~ PriorFinalGrade, data = .) %>%
    summ()

testScoresAll %>%
    lm(FinalMath ~ PredAvgiAFM + PriorFinalGrade, data = .) %>%
    summ()

testScoresAll %>%
    lm(FinalMath ~ int_iAFM + PriorFinalGrade, data = .) %>%
    summ()

testScoresAll %>%
    lm(FinalMath ~ int_iAFM_reverse + PriorFinalGrade, data = .) %>%
    summ()
```

	R^2	Adjusted R^2	log-likelihood	AIC	BIC
Model 1	0.43	0.43	-769.3078	1544.616	1554.904
Model 2	0.46	0.46	-762.7504	1533.501	1547.218
Model 3	0.5	0.5	-754.2367	1516.473	1530.191
Model 4	0.49	0.49	-756.2946	1520.589	1534.307

Areas to be improved

- Apply various feature transformation techniques to original dataset
    1. Normalize numerical features to eliminate the effect of outliers

```
FIXED EFFECTS:
-----------------------------------------
                      Est.   S.E.   z val.
------------------ ------ ------ --------
(Intercept)           1.57   0.04    36.47
opportunity0          0.09   0.00    24.82
-----------------------------------------
```

    2. Convert categorical features to one-hot representation
- Incorporate more relevant variables into model training
    1. Explore more features in Bernacki dataset (e.g.,)
    2. Identify important features using correlation analysis
    3. Feature engineering (e.g., correctness of previous problem)
- Choose proper metrics to evaluate the performance of models