

# Marist Universal Student Experience

Early detection, monitoring and intervention of academically at-risk students

Eitel J. M. Lauría<sup>1</sup>, Edward Presutti<sup>2</sup>, Maria Kapogiannis<sup>2</sup>

<sup>1</sup>School of Computer Science & Mathematics, Marist College, Poughkeepsie, NY, USA

<sup>2</sup>Data Science & Analytics Group, Marist College, Poughkeepsie, NY, USA

{Eitel.Lauria, Edward.Presutti, Maria.Kapogiannis}@marist.edu

IT

# Marist Universal Student Experience (MUSE)

- Collaborative Team
- History of Learning Analytics at Marist
  - Original Research
  - Operationalization
- How it Works, Creating the Unit of Analysis
- Stacked Ensemble
- Visualizing the Probabilities



\* Additional slides provided with a list of publications and additional faculty quotes

# Enterprise Computing, Marist Universal Student Experience (MUSE)

## ➤ The Team

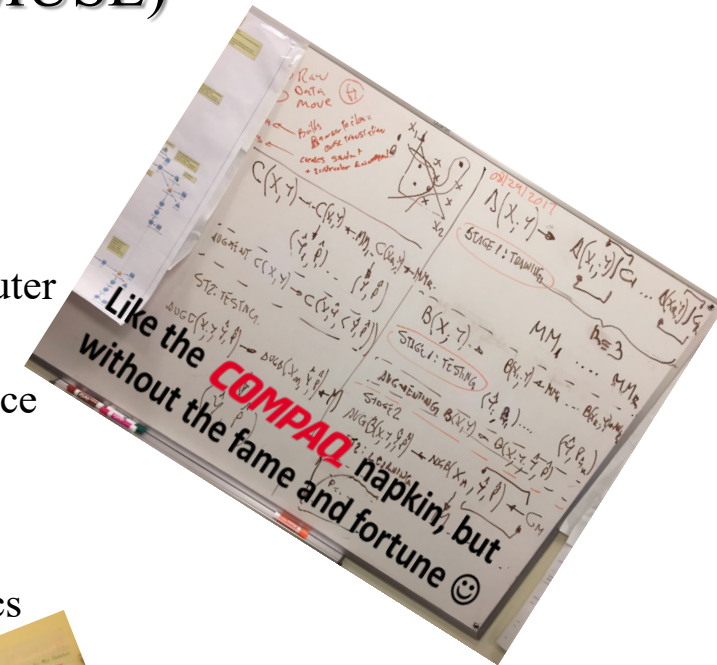
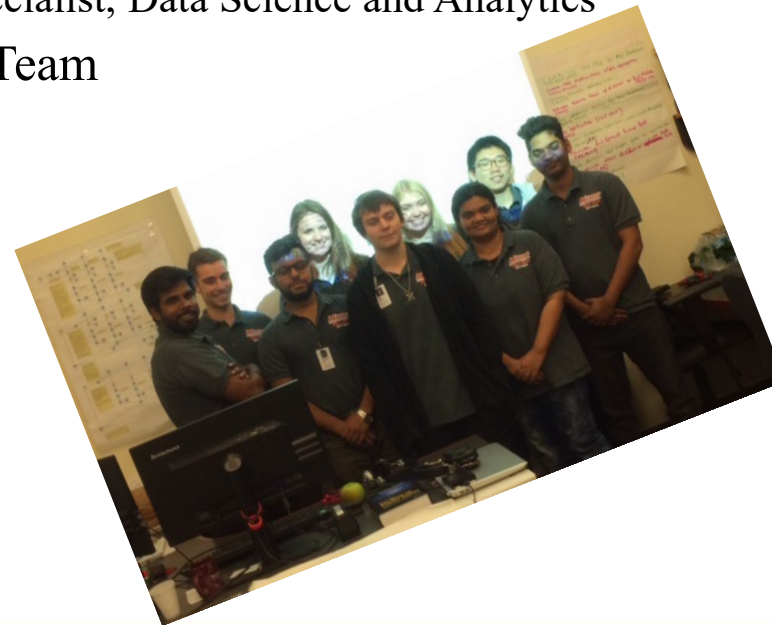
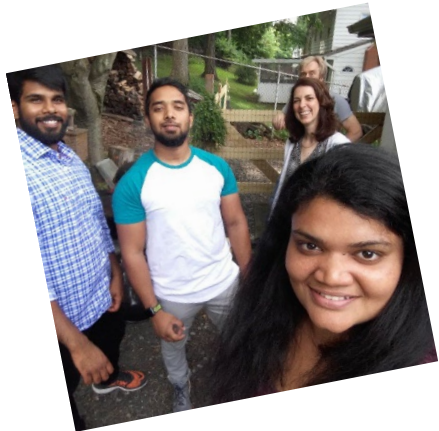
### ▪ Faculty Researchers

- ✓ Eitel J.M. Lauría, PhD, Professor and Graduate Director - School of Computer Science & Mathematics, Lead Data Scientist - Learning Analytics Initiative
- ✓ Erik Moody, PhD, Assistant Professor, School of Social & Behavioral Science

### ▪ Information Technology

- ✓ Edward M. Presutti, Assistant Director, Data Science and Analytics
- ✓ Maria T. Kapogiannis, Data Analytics Specialist, Data Science and Analytics
  - Original Student Implementation Team

- Anuya Kamath
- Stephen Bohner
- Erik Stenton
- Jillian Preece
- Yi Lin
- Kaushik Srinivasapuram
- Sumit Bansode





# Enterprise Computing, Marist Universal Student Experience (MUSE)

## Open Academic Analytics Initiative (2011 – 2013)

**Pilot Schools:**

Savannah State, North Carolina A&T, Cerritos College, College of the Redwoods

- EDUCAUSE Next Generation Learning Challenges (NGLC) grant
- Funded by Bill and Melinda Gates Foundation
- Create “early alert” framework based on open source tools
- Predict academically at-risk students in initial weeks of a course
- Computerworld Honors Laureate and Finalist in the Emerging Technology category
- Campus Technology Innovator Award
- Prototypes Originally Developed Using SPSS Modeler



## Learning Analytics Processor v1 (2013-2017)

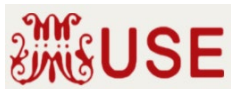
**Pilot Schools:**

University of Gloucestershire, Cardiff Metropolitan University, Aberystwyth University, University of Greenwich

- Early detection framework part of Aperio
- Chosen in late 2015 as a key component of the UK’s national analytics infrastructure provided by JISC
- Pilots at UK universities and NCSU during 2016 / 2017



## Learning Analytics Processor v2



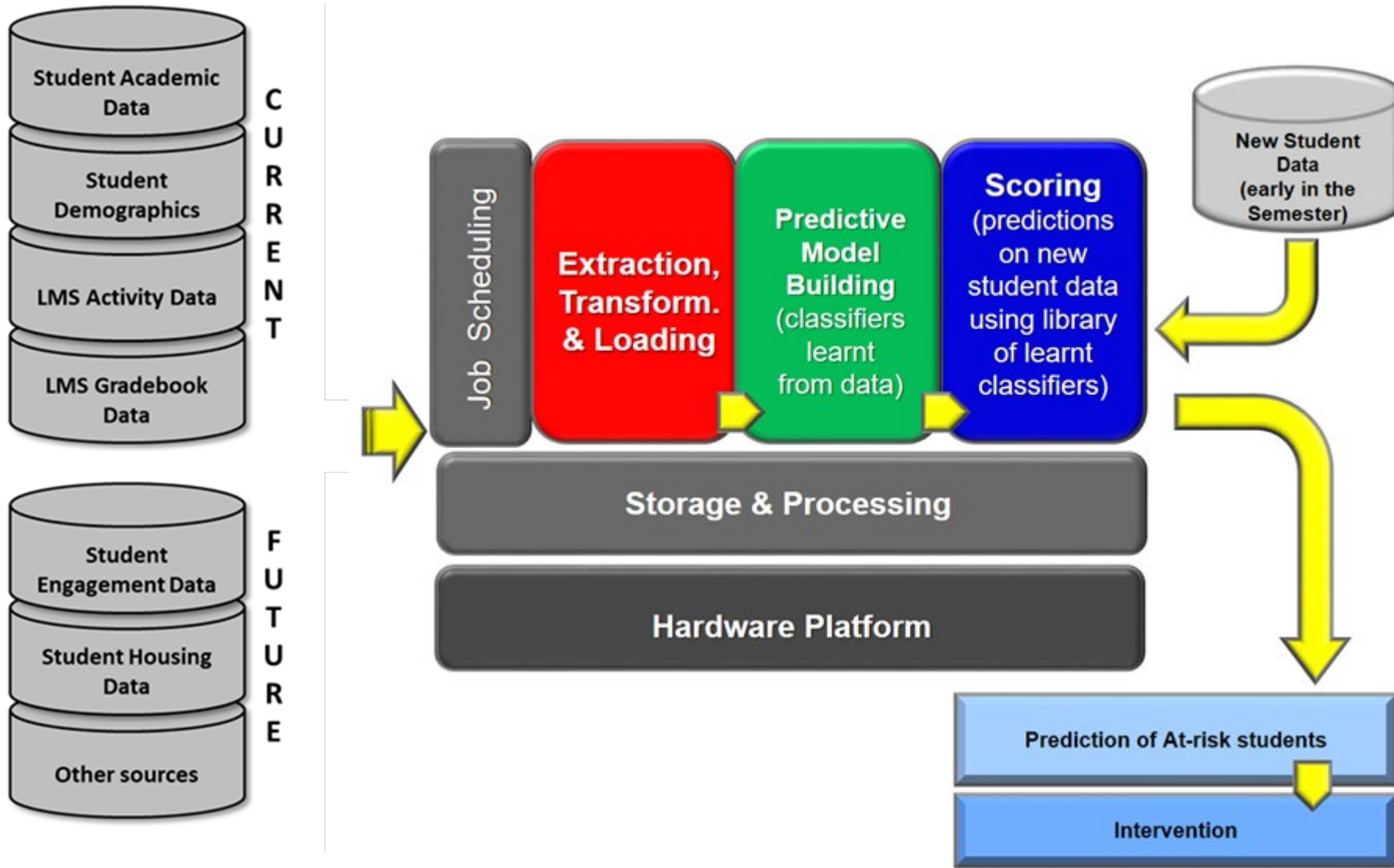
(2017-2019)

- New predictive analytics engine (based on a **stacked ensemble of classifiers**)
  - Enhanced ETL process
  - Web-based GUI and dashboard integrated into iLearn
  - Open source contribution to Aperio
- **MUSE, Marist Universal Student Experience**



# Enterprise Computing, Marist Universal Student Experience (MUSE)

## Analytics Process @ Marist

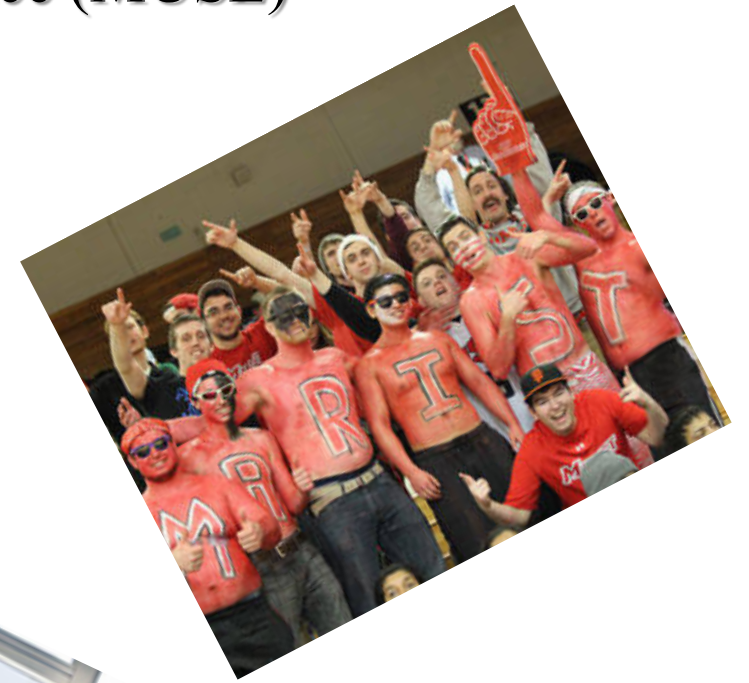


IT

# Enterprise Computing, Marist Universal Student Experience (MUSE)

## Input Data Features

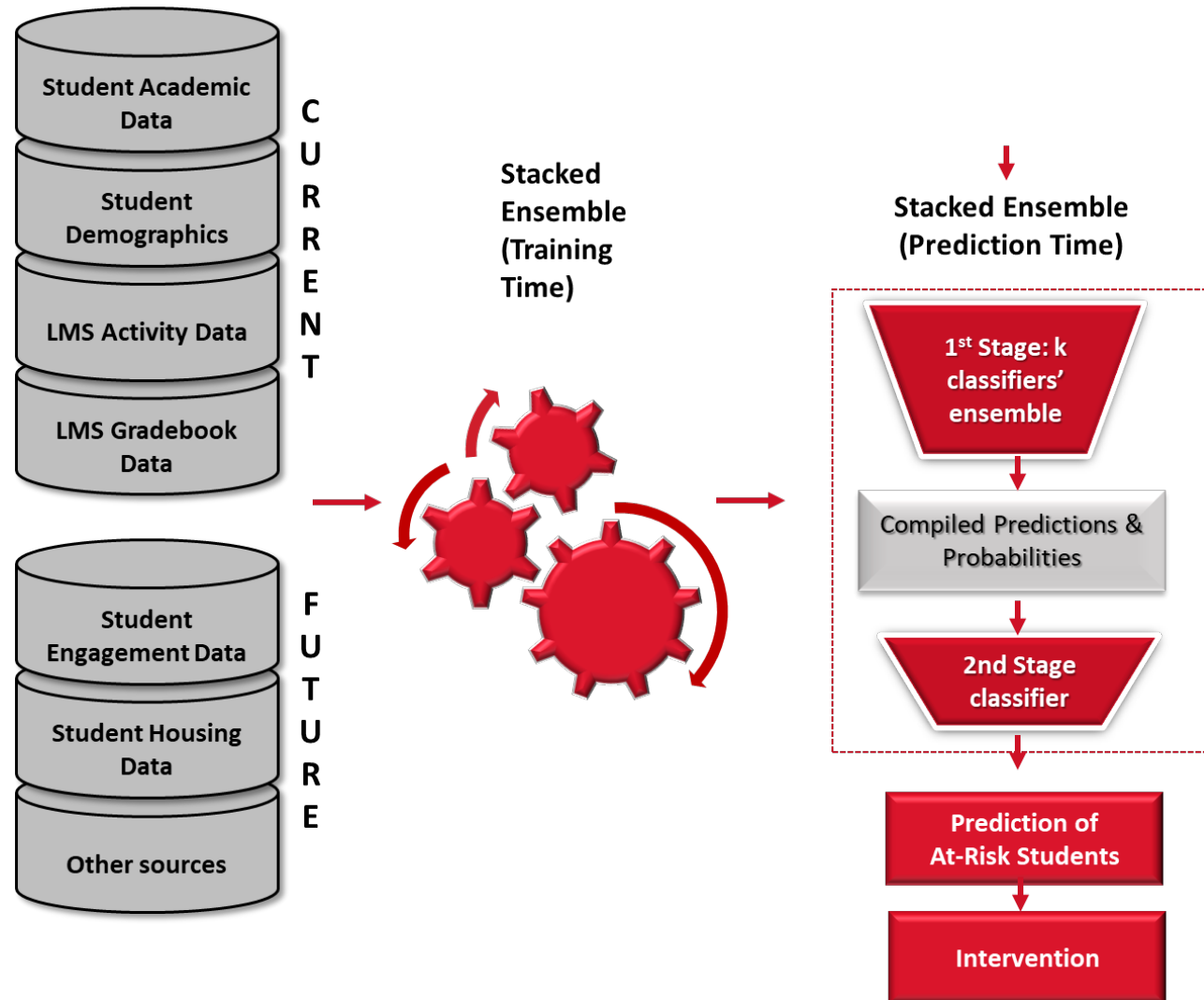
- Gender
- Age
- Class (Freshman, Sophomore, Jr, Sr.)
- Aptitude Score (e.g. SAT)
- Cumulative GPA
- Course Size
- Discipline (SCI, LA, CSM, BUS, SBS, CA)
- LMS Total Activity (weeks 1-6 + sum)
- LMS Login (weeks 1-6 + sum)
- LMS Content Read (weeks 1-6 + sum)
- Gradebook Composite Score (weeks 1-6)



IT

# How does it work ? (Stack Ensemble Architecture)

- Replace majority voting of traditional ensembles with statistical learning (learn from the predictions of a first stage ensemble)
- Introduced by Wolpert (1992)
- van der Laan et al. (2007) provided proof that a 'super learner' ensemble represents an asymptotically optimal system for learning.
- Lately stacked ensembles have been top performers of the Netflix Award and Kaggle competitions

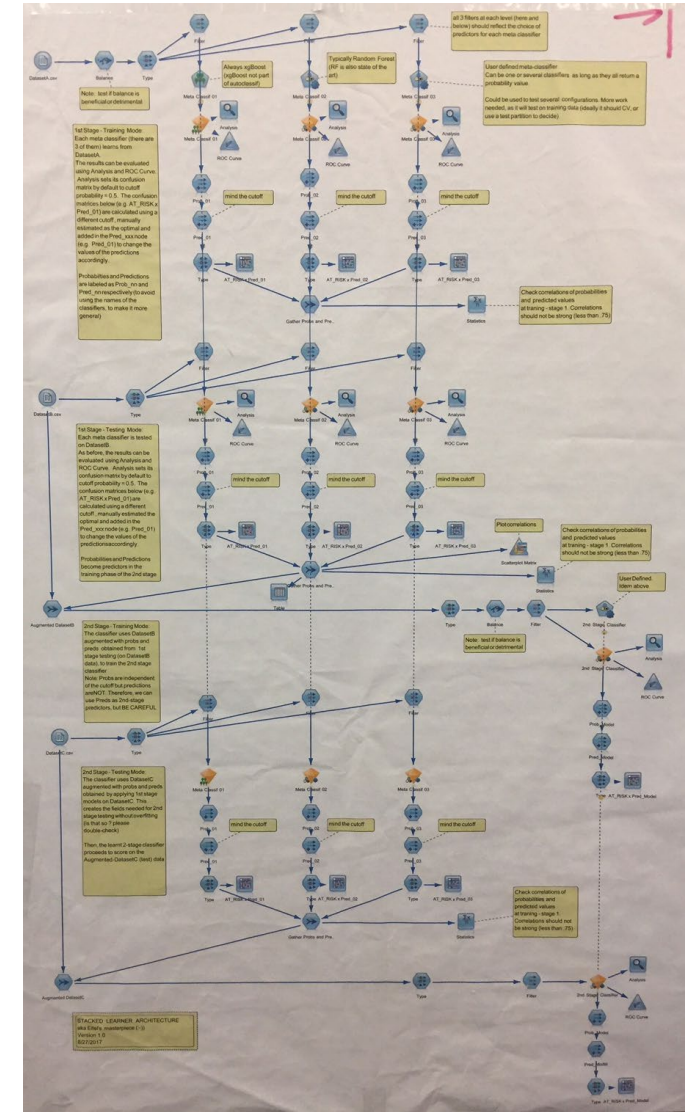




# Enterprise Computing, Marist Universal Student Experience (MUSE)

## Stacked Ensemble

- Training and testing a two-stage stack with  $k$  binary classifiers in the first stage, one binary classifier in the second stage, and 3 independent data sets A, B, C, to avoid data leakage.
- After the stack is trained, tuned and tested, it can be used to make predictions on new data D.
- Learning algorithms used to train base models should be as different as possible so that the predictions made by them have relatively low correlations ( $< 0.75 - 0.80$ ).
- If predictions are highly correlated, it indicates that the base models map very similar hypothesis functions, which defeats the purpose of using a stack.





# Enterprise Computing, Marist Universal Student Experience (MUSE)

## Classifiers

### First-stage Classifiers

Code	Description
RF	RF: The Random Forests algorithm (Breiman, 2001), a variation of bagging applied to decision trees
NN	NN: A feed-forward neural network (multilayer perceptron) with one hidden layer and varying number of units.
NB	The naïve Bayes algorithm with kernel estimation, to estimate the densities of numeric predictors.
XB	XGBtree (Chen and Guestrin, 2016), a recent implementation of the gradient boosted tree algorithm.

### Second-stage Classifiers

Code	Description
LOG	Regularized logistic regression using the LibLinear library (Fan et al., 2008)
LMT	Logistic model trees (Landwehr et al., 2005)

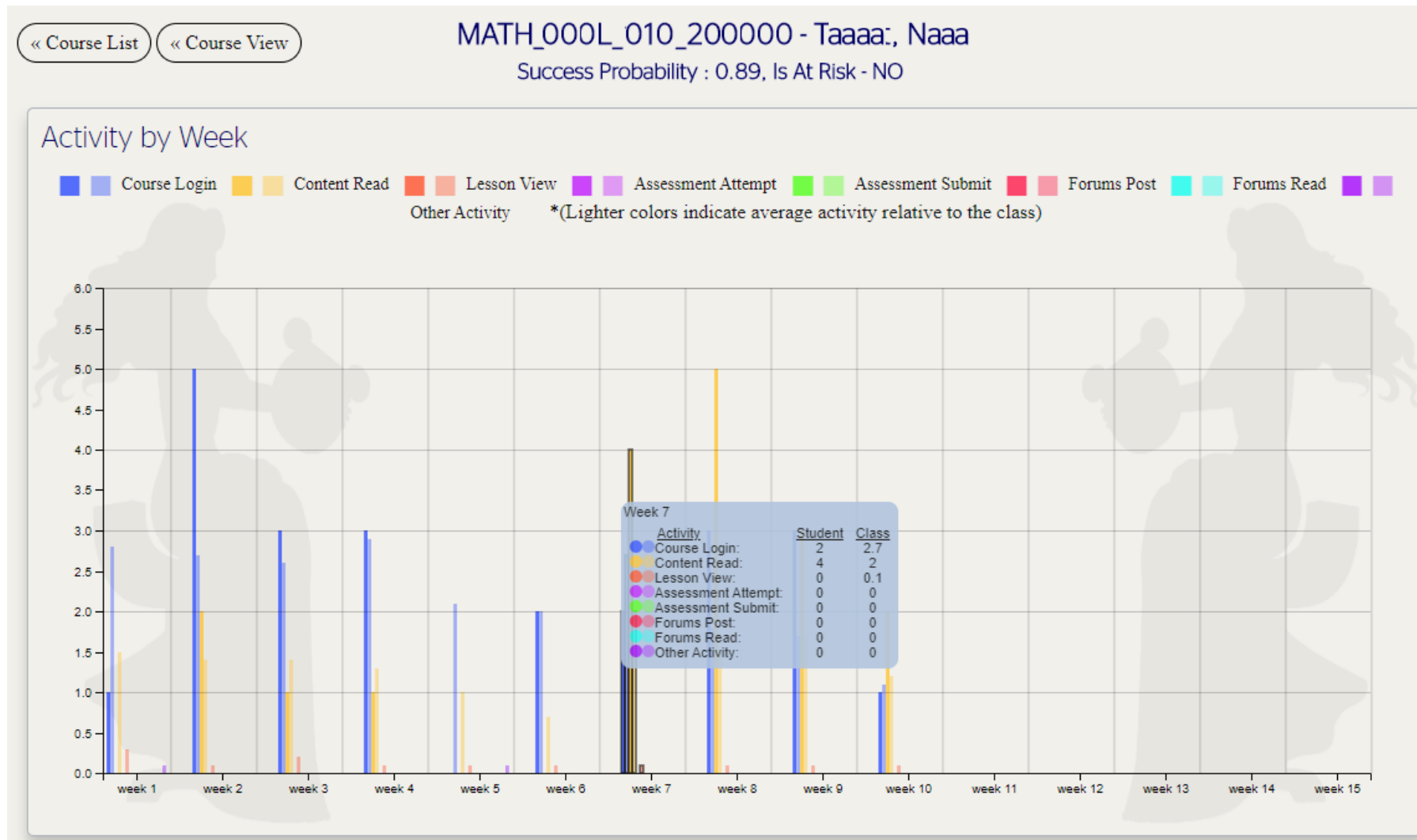
#### Relevant References:

- Breiman, L., 2001. Random Forests. Mach Learn 45, 5–32.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. CoRR abs/1603.02754.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: A Library for Large Linear Classification. J Mach Learn Res 9, 1871–1874.

The logo consists of the letters 'IT' in a white, serif font, centered within a red semi-circular shape that is part of a larger graphic element on the right side of the slide.

# Enterprise Computing, Marist Universal Student Experience (MUSE)

Course Activity



IT

# Enterprise Computing, Marist Universal Student Experience (MUSE)

Course Probabilities Active

« Course List

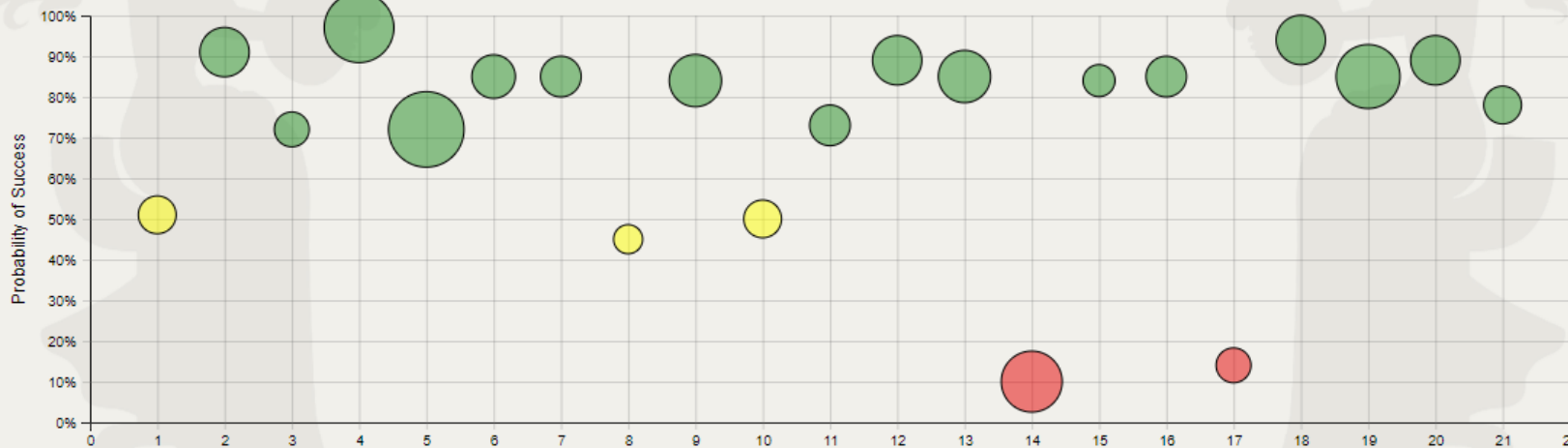
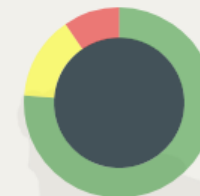
## Class Summary - MATH\_000L\_000\_200000

Class Start Date	Class End Date	Enrollments	At Risk Count	Max Activity	Mean Activity	Min Activity
2000-01-16	2000-05-11	21	2	34	1.86	1

Probability of Success - In the presence of insufficient data, the system displays the GPA if it is present in the student record, scaled between 0% - 100%

■ At Risk
 ■ Possibly at Risk
 ■ Not at Risk
 +   Insufficient data for prediction

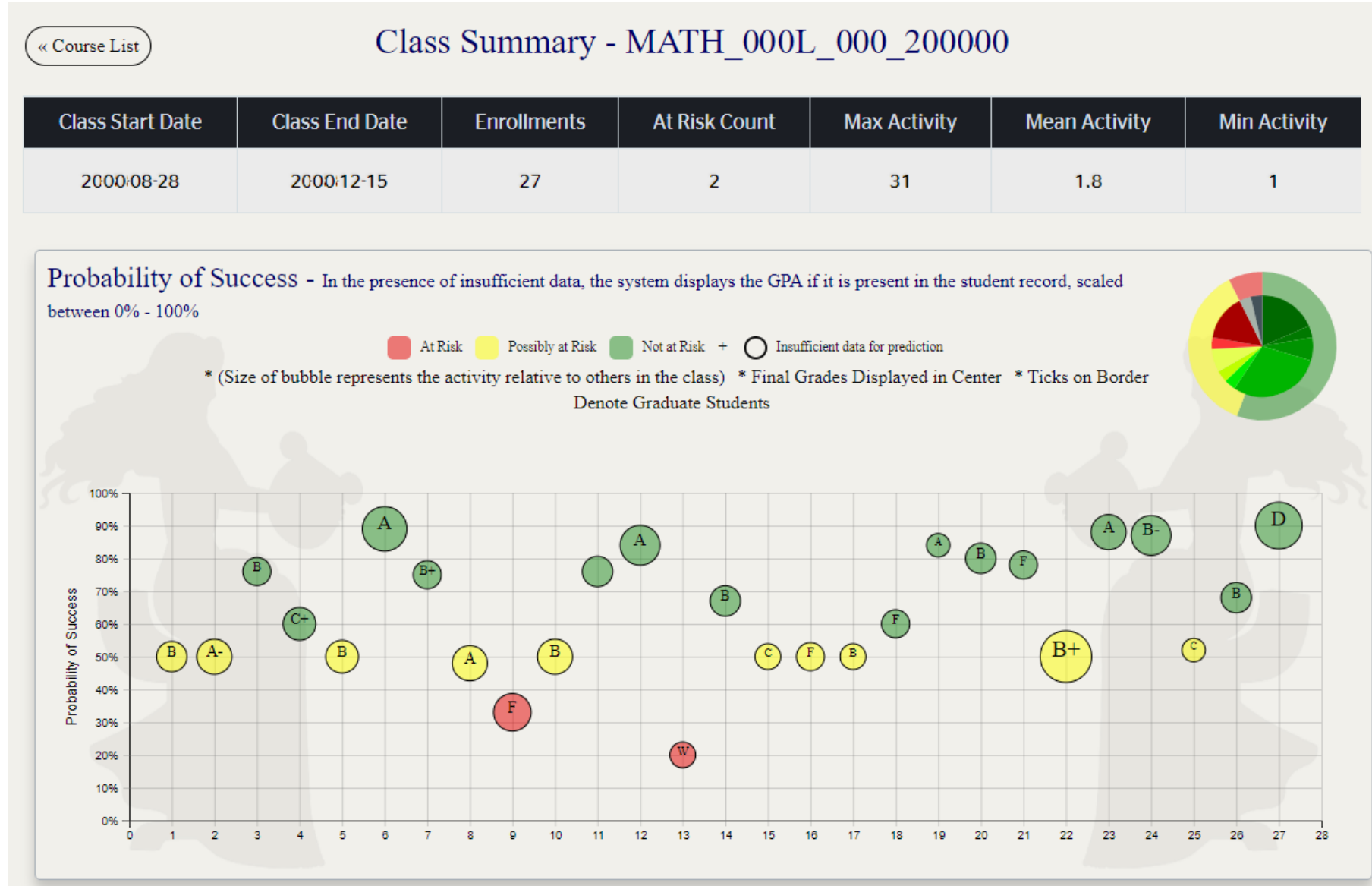
\* (Size of bubble represents the activity relative to others in the class) \* Final Grades Displayed in Center \* Ticks on Border Denote Graduate Students



IT



# Enterprise Computing, Marist Universal Student Experience (MUSE)



## Enterprise Computing, Marist Universal Student Experience (MUSE)

- ✓ ***“I couldn't wait to open up Muse for a brief glance. Just to let you know, I was quite pleased and stunned at the accuracy thus far. The students identified are the same students I just reached out to via email this a.m. requesting that they make appointments to see me re: their performance. Thank you!”***
- ✓ ***“This is a great tool, I really like the activity size on the chart, very informative without having to click on each student to try to determine engagement. “***
- ✓ ***“I was amazed when I opened the assessment early on and it was extremely helpful to have my initial thinking affirmed. I was able to reach out to the students and am hopeful that their grades will improve. Thank you.”***

## Enterprise Computing, Marist Universal Student Experience (MUSE)

- Current choice of classifiers is discretionary (state of the art classifiers that yield probabilities and different enough to cover as much as possible of the hypothesis space)
- MUSE uses a fixed 2-stage stack with 3 base classifiers
- Nonetheless, it provides first-time insight of the use of a stacked ensemble architecture in the domain of learning analytics and early detection of academically at-risk students





Questions?

# Enterprise Computing, Marist Universal Student Experience (MUSE)

- Lauría E., Presutti E, Kapogiannis M., “Of Stacks and Muses: Adventures in Learning Analytics at Marist College” (2019), Poster presented at LXAI@ICML 2019. LXAI Research workshop co-located with the Thirty-sixth International Conference on Machine Learning, Long Beach, CA, June 10, 2019 (upcoming)
- Lauría E., Presutti E, Kapogiannis M. (2019), “MUSE: Marist Universal Student Experience”, 2019 Enterprise Computing Community Conference, Marist College, June 9-11, 2019 (upcoming)
- Moody E., Lauría E., Presutti E., The development of an academic early alert system at Marist College., Online Learning 2018: Assessment, Analytics and Student Success. The Westin Harbour Castle, Toronto, Ontario October 16 -18, 2018. <https://onlinelearning.secure-platform.com/a/solicitations/1/sessiongallery/54/application/2132>
- Lauría E., Presutti, E., “MUSE: Marist Universal Student Experience”, Apereo Lightning Talks: Analytics Theme, March 28, 2018, <https://www.apereo.org/content/apereo-lightning-talks-analytics-theme>
- Lauría E., Presutti E, Kapogiannis M., Kamath A., “Stacking Classifiers for Early Detection of Students at Risk”, Proceedings of the 10th International Conference of Computer Supported Education (CSEDU 2018), Funchal Madeira, Portugal, March 15-17, 2018, vol 1, 390-397
- Lauría E., Presutti, E., “MUSE: Marist Universal Student Experience”, 10th International Conference of Computer Supported Education (CSEDU 2018), Funchal Madeira, Portugal, March 15-17, 2018
- Lauría E., Presutti E., Sokoloff M., Guarino, M., "Crossing the Chasm to Big Data: Early Detection of at-Risk Students in a Cluster Computing Environment", Practitioner Track Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK'17). Simon Fraser University, Vancouver, Canada (13–17 March)
- Lauría E., Kuck P., Presutti E., Jayaprakash S., Sokoloff M., "Learning Analytics at Marist College: From a Single Node Prototype to a Cluster Computing Platform", 2016 Enterprise Computing Community Conference, Marist College, June 12-14, 2016
- Lauría, E. “Early detection of academically at-risk students”, IBM’s Apache Spark Maker Community Event, Galvanize, San Francisco, June 6, 2016.
- Jayaprakash S., Baron J., Gilbert G., Lauría E., Moody E., Reagan J. "Open-source Academic Early Alert and Risk Assessment API", LAK 2015, Practitioner’s Track
- Bainbridge J., Melitski J., Zahradnik A., Lauría E., Jayaprakash S., Baron J., "Using learning analytics to predict at-risk students in graduate online public affairs and administration education", Journal of Public Affairs Education, 21(2), Spring 2015, 247-
- Jayaprakash S., Moody E., Lauría E., Regan J., Baron J., "Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative", Journal of Learning Analytics, 1(1), 2014, 6-47 (available online [here](#))
- Jayaprakash S., Lauría E., (2014), "Open Academic Early Alert System: Technical Demonstration", Proceedings of LAK 14 (Learning Analytics and Knowledge Conference) , Indianapolis, IN, March 24-28, 2014
- Regan, J., Moody, E., Lauría, E., Jayaprakash, S., Jonnalagadda, N., Baron, J. "Learning Analytics: Coming to a Classroom Near You". Poster presented at the Annual Meeting of the Association for Psychological Science, Washington, D.C., May 2013
- Lauría E., Moody E. , Jayaprakash S., Jonnalagadda N., Baron J. (2013), "Open Academic Analytics Initiative: Initial Research Findings", Proceedings of LAK 2013 (Learning Analytics and Knowledge Conference) , Leuven, Belgium, April 8 - 12, 2013
- Lauría E., Baron J., Devireddy M., Sundararaju V., Jayaprakash S. (2012), "Mining academic data to improve college student retention: An open source perspective", Proceedings of LAK 2012 (Learning Analytics and Knowledge Conference) , Vancouver, BC, Canada, April 29 - May 2, 2012
- Lauría E. , “Mining Sakai to Measure Student Performance: Opportunities and Challenges in Academic Analytics” (presentation of the Open Academic Analytics Initiative and discussion of preliminary results), IBM Academy of Technology Affiliates, October 4, 2011
- Lauría E. , Baron E. , “Mining Sakai to Measure Student Performance: Opportunities and Challenges in Academic Analytics”, 2011 Enterprise Computing Community Conference, Marist College, Poughkeepsie, NY, June 12-14 2011

A large red semi-circle graphic on the right side of the slide, containing the white letters "IT" in a serif font.

# Enterprise Computing, Marist Universal Student Experience (MUSE)

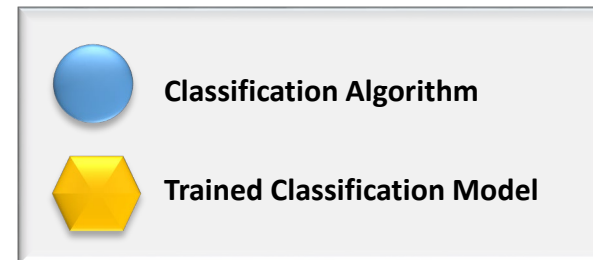
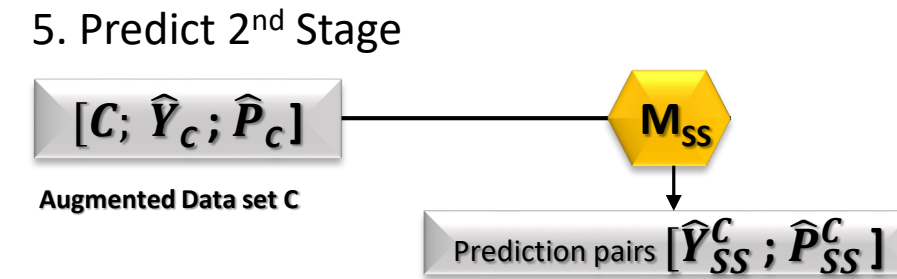
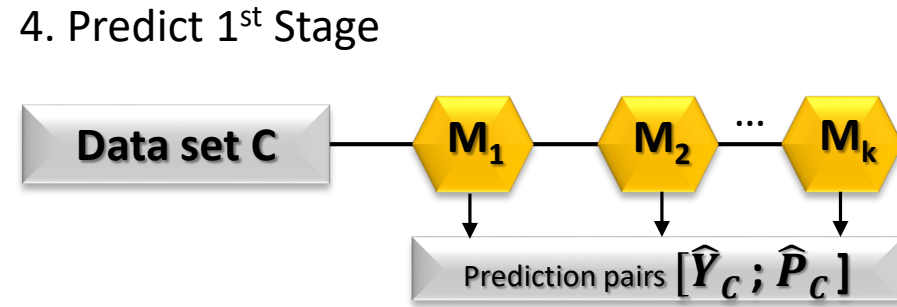
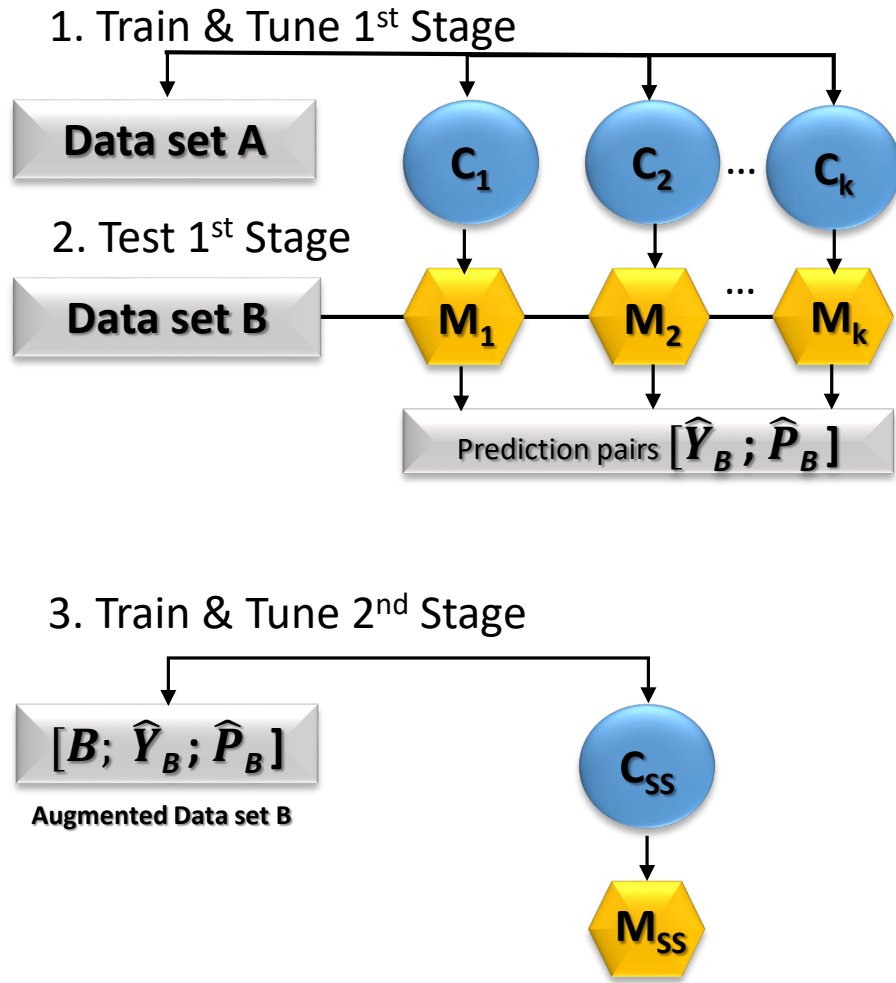
- ✓ *"I couldn't wait to open up Muse for a brief glance. Just to let you know, I was quite pleased and stunned at the accuracy thus far. The students identified are the same students I just reached out to via email this a.m. requesting that they make appointments to see me re: their performance. Thank you!"*
- ✓ *"The dashboard for each class instantly giving me a bubble chart view of my students is a great way to very quickly see a summary of where everyone is in the class .. hovering over the students who are at risk to identify, and confirm my own belief instantly is powerful, but it's more powerful to hover over a student at risk and be like "huh... okay, that's interesting... now, what do I do about that!"*
- ✓ *"This is a great tool, I really like the activity size on the chart, very informative without having to click on each student to try to determine engagement. "*
- ✓ *"I was amazed when I opened the assessment early on and it was extremely helpful to have my initial thinking affirmed. I was able to reach out to the students and am hopeful tha their grades will improve. Thank you."*
- ✓ *"This is very useful information. I handled my courses (301/302) differently, i.e., I did not reach out early to the students identified in my 302 classes, but did reach out to the students in my 301. I do credit the improved grades of the 301 students to the early intervention. I chose to run my own micro-mini field experiment with 301 as the students in 302 will have much more opportunity to improve their grades in the second half of the semester due to a research paper. The success of the paper will depend in part on their use of the resources posted in iLearn. So, this information provides me with a grand opportunity to emphasize the importance of utilizing iLearn. Thank you again for the opportunity"*
- ✓ *"I emailed the students and scheduled a time to meet. Due to my class being online, students don't always ask to meet but sometimes they just need additional explanation. I also worked on creating a more structured schedule for certain students who had red bubbles"*
- ✓ *"Depending upon the situation, i.e., I am aware of some health issues in a number of the students early on, I have either filed an Academic Warning and requested a meeting or simply requested a meeting."*
- ✓ *"Have been meeting with the at risk students, at least the ones that have hope of passing the class"*



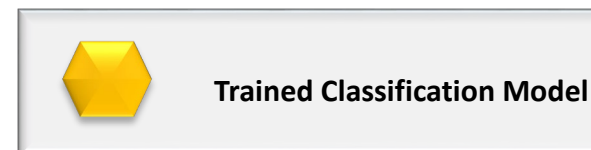
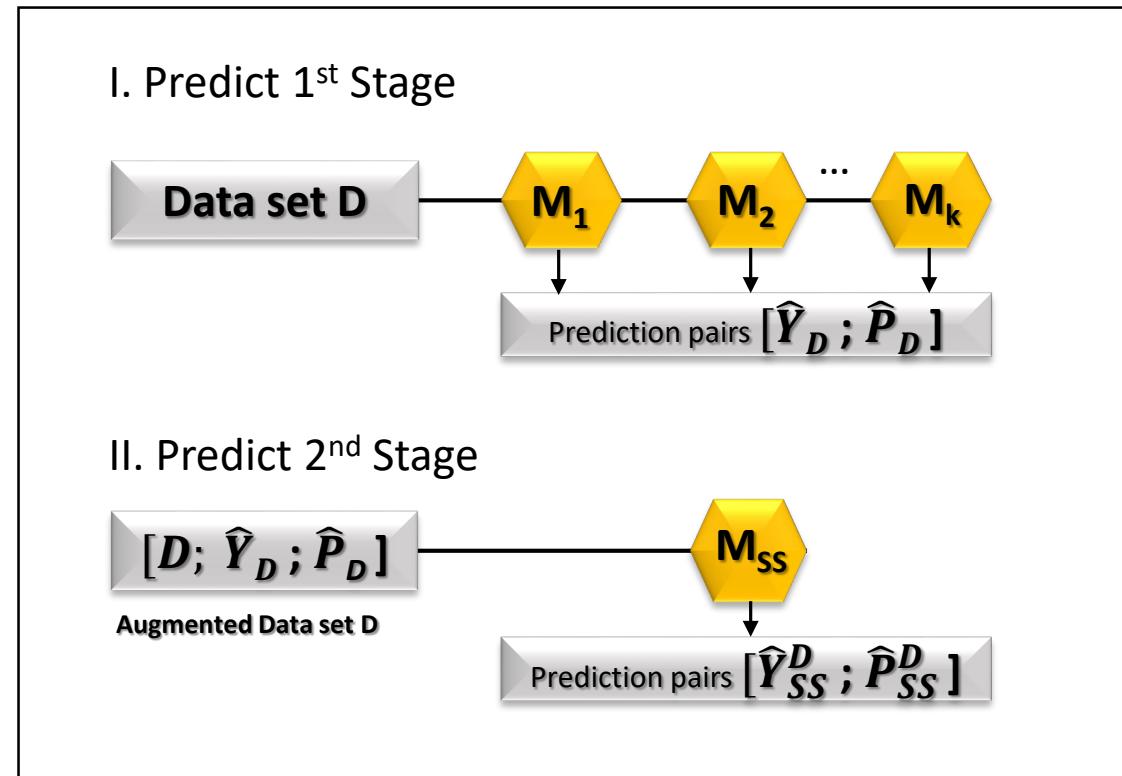
# Additional Slides

IT

# Stack Architecture: Model Building



# Stack Architecture: Model Scoring



IT