# Broader and Earlier Access to Machine Learning

Thomas Way, Lillian Cassel, Paula Matuszek, Mary-Angela Papalaskari,
Divya Bonagiri and Aravinda Gaddam

VILLANOVA UNIVERSITY

Department of Computing Sciences

## Goal

Create educator modules for teachers in any discipline to include relevant Machine Learning concepts.

- Identify Relevant Topics
- Produce Ready Modules
- Disseminate Online

## Module Example

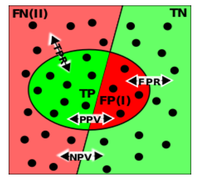
Complete material to teach machine learning in context, understandable by non-computer scientists. Typical module contains:

- **Instructor Overview** - background, concepts
- **Handouts** - Activities
- **Data** - general and discipline specific
- **Evaluation** - pre/post tests, quizzes

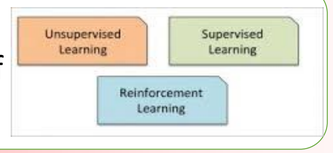
## Modules

***Intro***
Fundamentals, background, Animal Game


***Evaluating Classifiers***
How to compare results



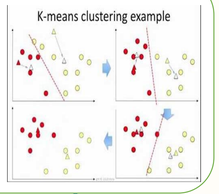

***Kinds of ML***
Overview of ML areas



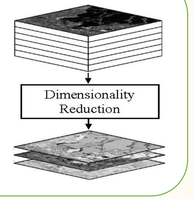

***Decision Trees***
Uses WEKA and data sets

***Text Classification***
introduces WEKA, classify tweets, authors WEKA


***K - Means***
Uses WEKA, advanced topic



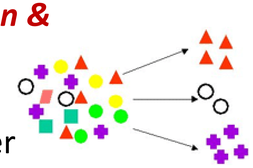

***Dimensionality Reduction***
Uses WEKA, advanced topic



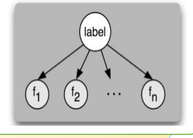

***Choosing Inputs***
Approaches to data sets

***Neural Networks***
Uses SimBrain software

***Classification & Clustering***
Hands-on, no computer needed

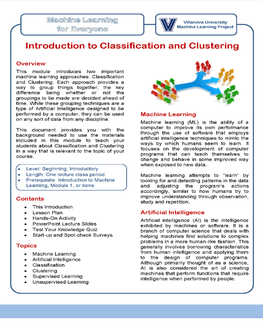


***Naive Bayes***
Uses Python or WEKA

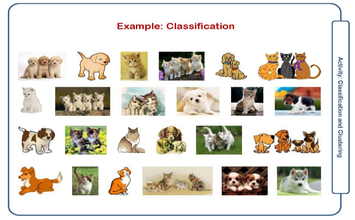



## Module Example: Classification & Clustering

Instructor Overview

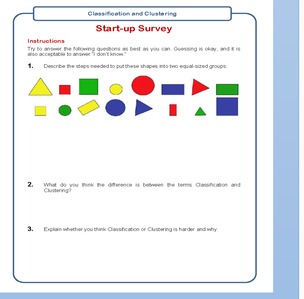


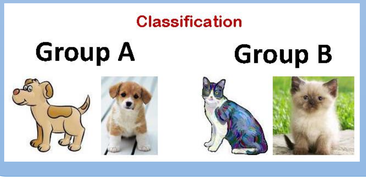

Hands-on Activity




Create groups using example groups or by finding similarities

Pre & Post Test





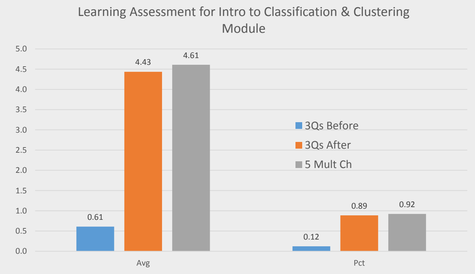

## Machine Learning

Computer programs that find patterns in data, enabling them to "learn" and make decisions based on that learning.

## Results

Results of pre/post tests for Classification & Clustering module show strong learning and good retention.

Pre test given **before** any material. Post test given **2 days later**.




41 students pre/post test scores went from low (12%) to solid (89%) understanding on identical questions, with strong ability to apply knowledge (92%) to similar problems.

## Future Plans

- Complete model design
- Gather more domain-specific data sets
- Disseminate via: **ComputingPortal.org/MachineLearning**

## Acknowledgements

Project funded in part by NSF DUE award 1141033. Thanks to Carol Weiss for learning measure design and evaluation.

21st Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE)
11th – 13th July 2016, Arequipa, Peru