Learning Objectives

Slide 1

In this section, I would like to discuss the skills you will develop throughout the semester.

We will start by learning the basics of R programming. This will give you the skills needed for performing more advanced data analysis as we progress through the semester.

We will then move to data analysis and visualization using the popular tidyverse R package.

Once we have a solid foundation in data analysis, we will focus on concepts from applied statistics and probability. This will help us understand the results we obtain when fitting machine learning models and effectively communicate their business value.

Next, we will spend a significant portion of the semester learning how to build machine learning models with the tidymodels package.

And finally, we will work on an analytics project that combines all the topics we have learned throughout the semester. This will give students a chance to develop their skills in communicating the business value discovered from analyzing data and predictive modeling.

Slide 2

We will be using the popular IDE, RStudio for our programming tasks this semester.

This semester, we will be learning about R data structures, such lists and data frames, how to write custom functions for data analysis tasks, and conducting an analytics project using a real-world data set.

Slide 3

Before diving into how to fit machine learning models, we will spend a lot of time learning the basics of data preparation and analysis with the tidyverse package.

Cleaning raw data and exploring it for insights is a prerequisite for most analytics projects. In fact, many data scientists report spending more than half their time cleaning data!

We will start by learning how to manipulate data in R to create tables and summary statistics.

Then, we will learn how to reshape data. This is an important skill to learn because it will enable us to take raw data and transform it to a format suitable for machine learning.

And finally, we will learn how to create meaningful data visualizations to aid in communicating our insights from a data analysis.

Slide 4

We will spend most of our time this semester studying different machine learning algorithms and how to implement them with the tidymodels package.

We will focus on two broad categories within machine learning.

Regression, where we are predicting numeric outcomes such as the selling price of a home, and Classification, where we are predicting categorical outcomes as “Yes”/”No” or “Will Purchase”/”Will Not Purchase”.

These will be explained in detail in the next video.