



Machine Learning Final Project
Logistic Regression Model
**Predicting Video Game
Success using ML**

Introduction

- Video games are a multi-billion dollar industry
 - Predicting success helps developers and publishers
 - Machine Learning enables data-driven decisions



Project Vision

- Use ML to predict whether a game will be successful
 - Leverage historical video game data
 - Provide insights for future game development



Problem Statement

- Game success is uncertain and risky
 - Traditional methods rely on intuition
 - Need a predictive ML-based solution



Objectives

- Analyze video game dataset
 - Preprocess and explore data
 - Train ML model to predict success



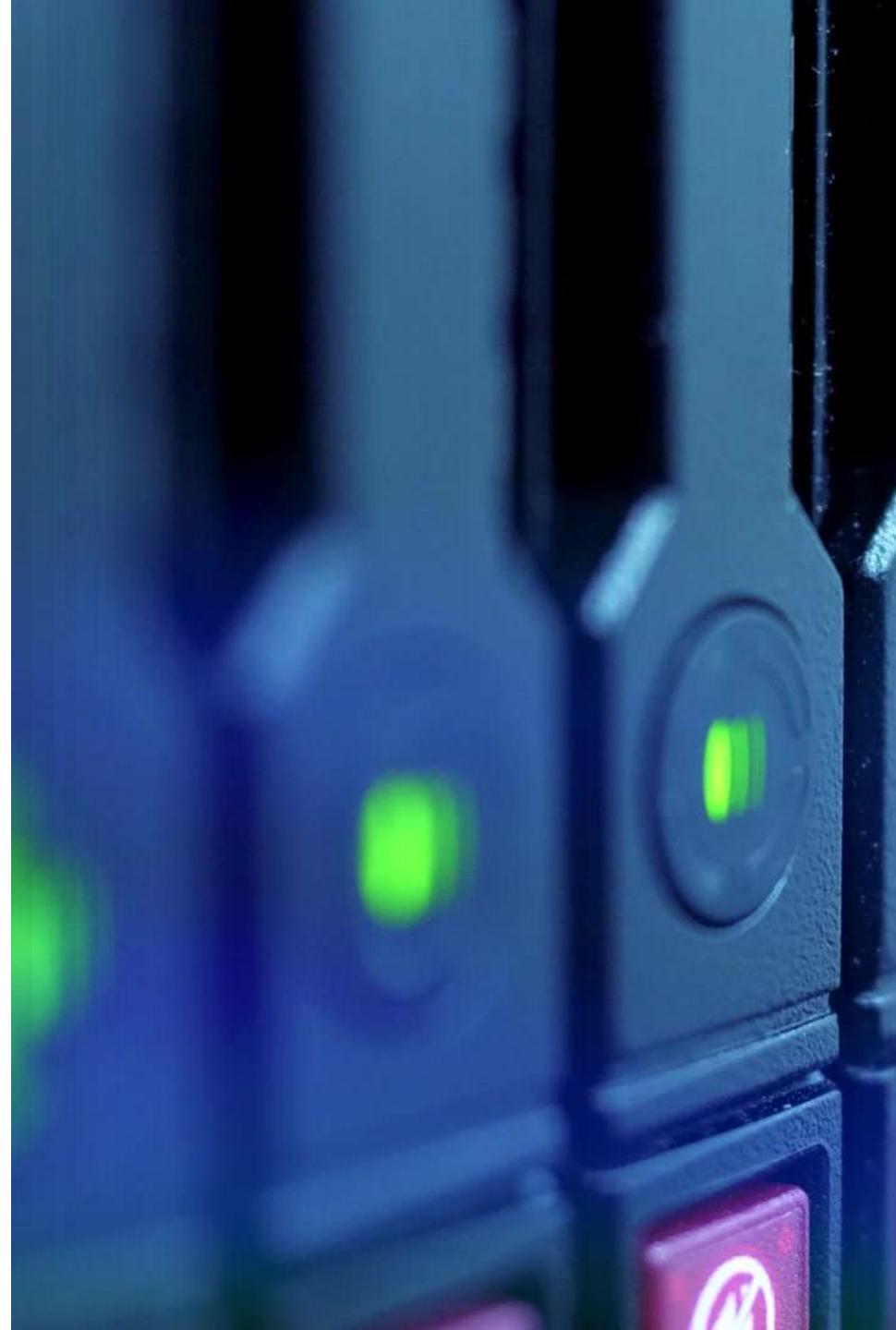
Project Scope

- Binary classification problem
 - Focus on structured numerical & categorical data
 - Logistic Regression model



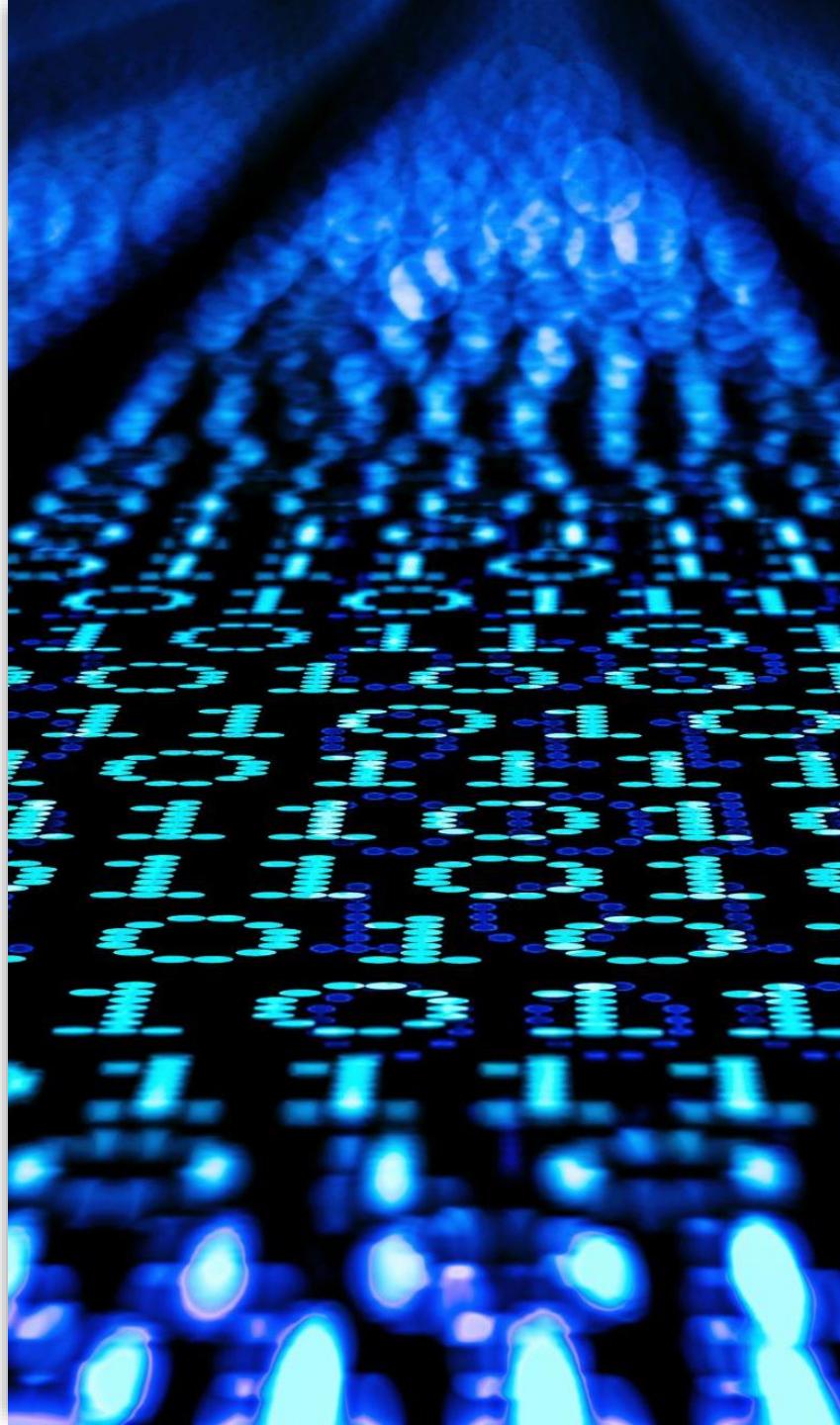
Dataset Overview

- Source: Video Game Sales Dataset
 - Features include sales, ratings, platform
 - Target variable: Success (Hit / Not Hit)



Data Preprocessing

- Handled missing values
 - Encoded categorical features
 - Feature scaling applied



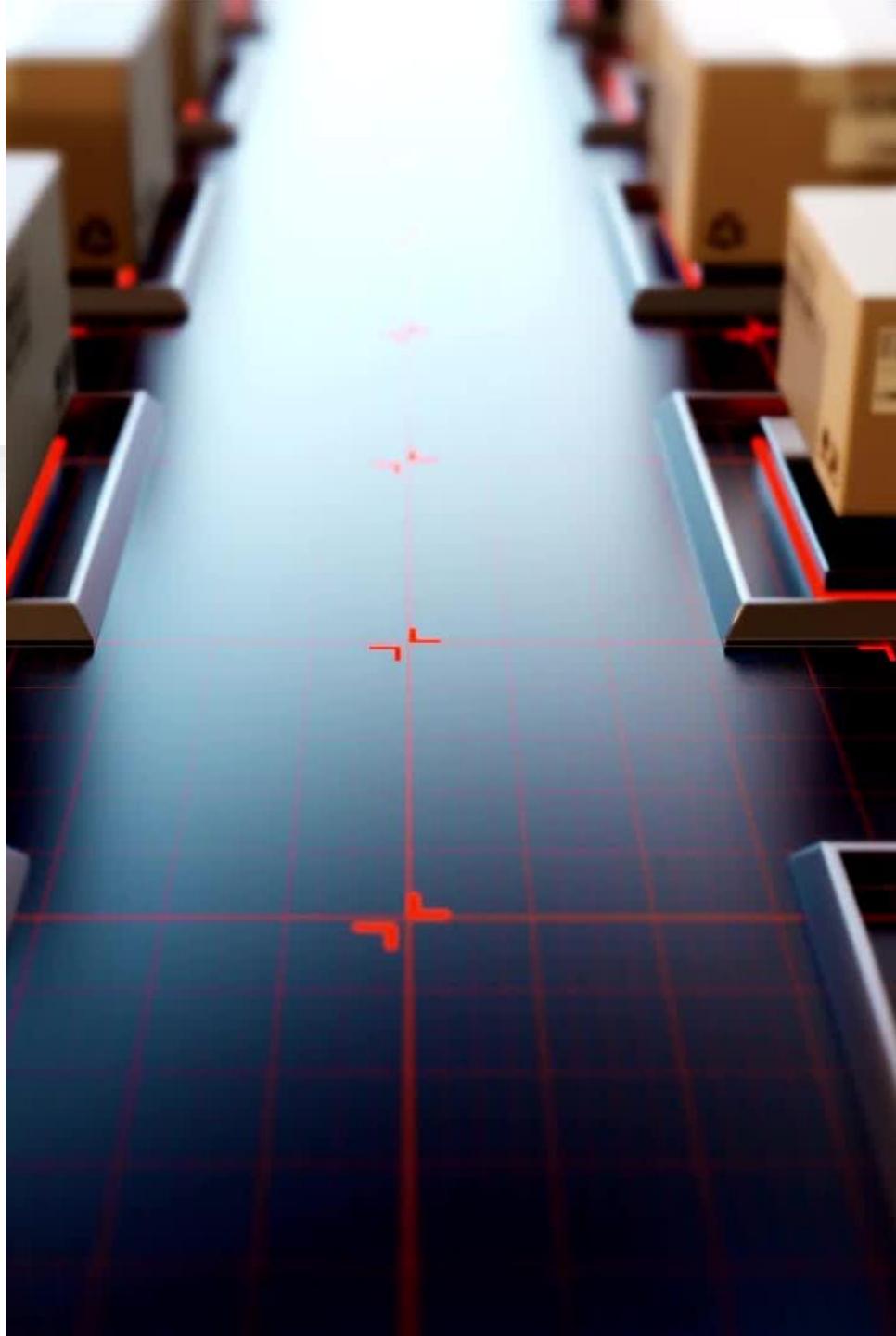
Exploratory Data Analysis

- Sales distribution analysis
 - Platform vs success trends
 - Correlation between features



ML Approach

- Logistic Regression chosen
 - Simple, interpretable, efficient
 - Well-suited for binary classification



Model Training

- Data split into training & testing sets
 - Model trained on training data
 - Evaluated using test data



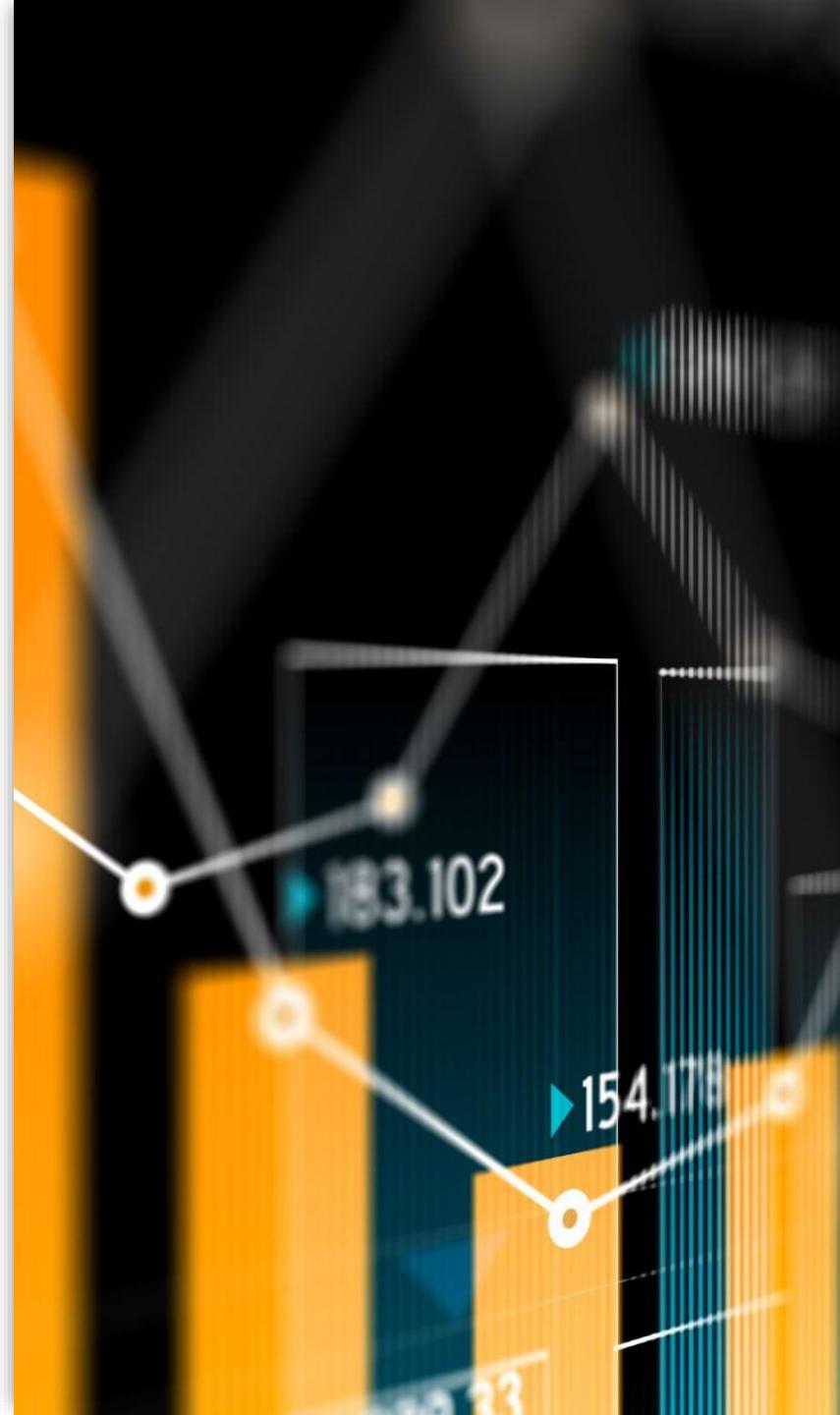
Accuracy Score

- Accuracy measures overall correctness
 - Model achieved strong performance
 - Indicates reliable predictions



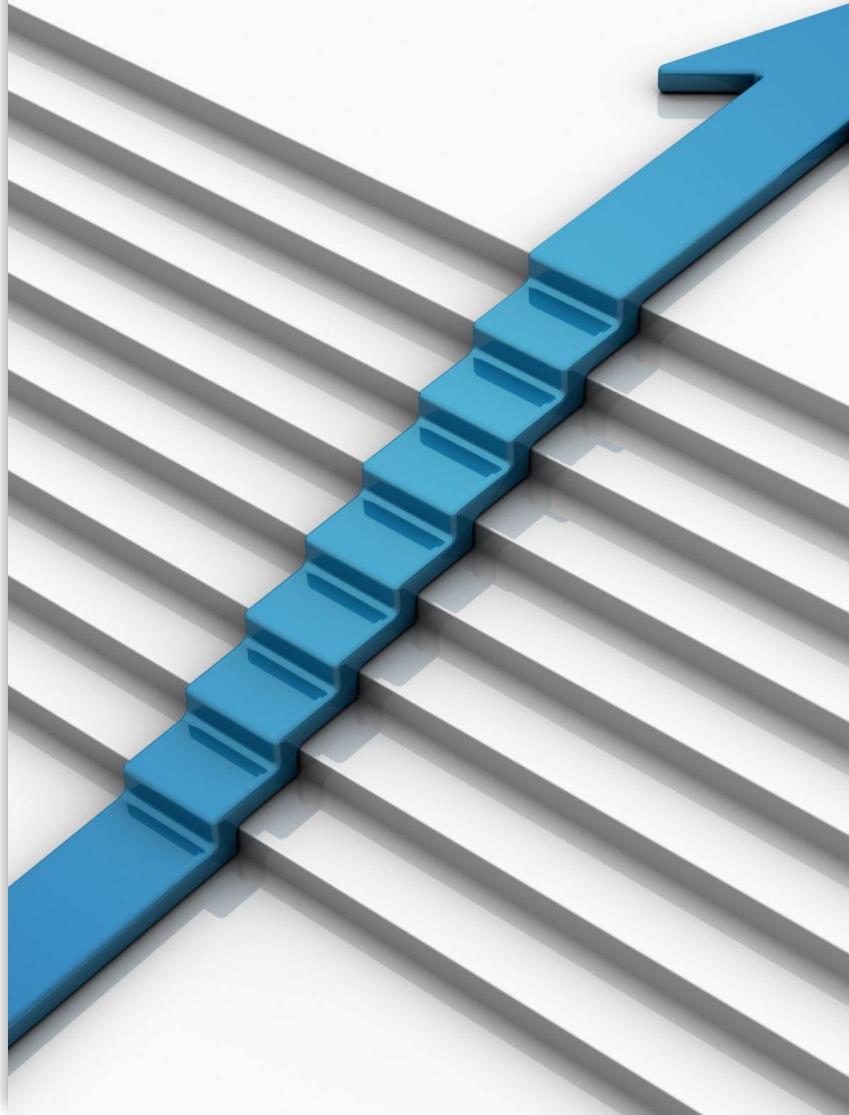
Confusion Matrix

- Shows TP, FP, TN, FN
 - Helps understand classification errors
 - Balanced performance observed



ROC Curve

- Plots True Positive Rate vs False Positive Rate
 - Evaluates model discrimination ability



ROC-AUC Score

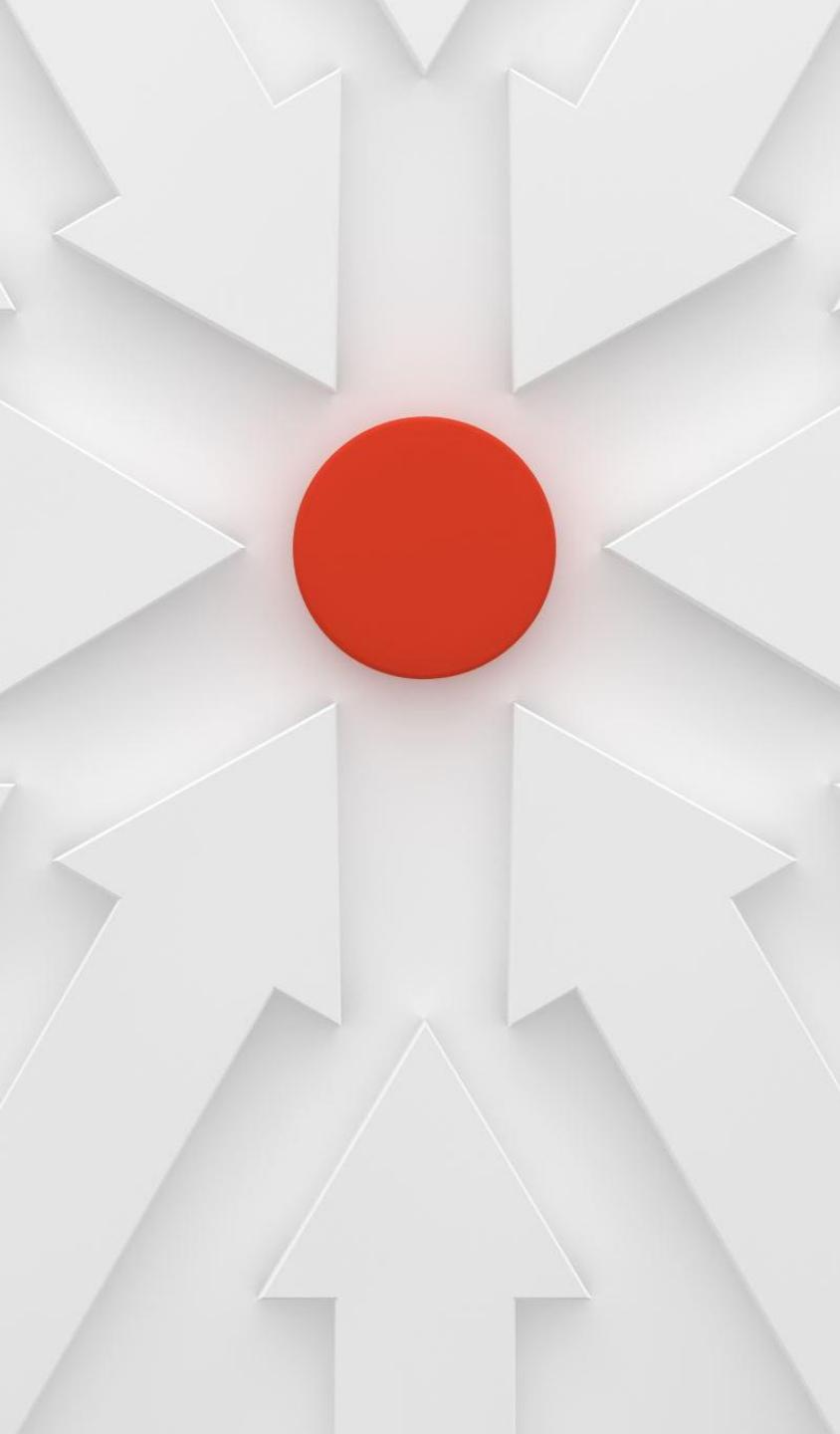
- Area Under ROC Curve
 - Higher AUC = better model
 - Model shows good separability

`predict_proba()` Explanation

- Returns probability estimates
 - Useful for confidence-based decisions
 - Not just class labels

The image shows a chalkboard with several mathematical derivations:

- A graph with a curve labeled $y = g(x)$ and a tangent line labeled $f(x)$.
- The text "Second Lines" is written next to the graph.
- A derivative definition: $f(x) = \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h}$
- An expansion of the numerator: $= \lim_{h \rightarrow 0} (x^2 + 2xh + h^2 - x^2) / h$
- Simplification steps:
 - $= \lim_{h \rightarrow 0} 2xh / h$
 - $= \lim_{h \rightarrow 0} h(2x + h)$
 - $= \lim_{h \rightarrow 0} h(2x + h) / h$

A large, abstract graphic occupies the left third of the slide. It features a central red circle surrounded by numerous white, three-dimensional arrows pointing towards it from all directions, creating a sense of convergence or focus.

Results & Insights

- Certain features strongly influence success
 - ML can assist decision-making
 - Model generalizes well



Conclusion & Future Work

- ML effectively predicts game success
 - Future: try advanced models
 - Include more features & data