A low-angle, black and white photograph of several modern skyscrapers reaching towards a cloudy sky. The perspective creates a sense of height and scale. A large, solid blue rectangle is overlaid on the left side of the image, containing the title text in white. White geometric lines, including a large 'L' shape and a diagonal line, are visible in the top-left and bottom-right corners of the image.

# MACHINE LEARNING IN REAL ESTATE

Giulia Brambilla - January 2020



# TABLE OF CONTENTS

## **1 Business Case**

Introduction

## **2 Regression**

AI techniques to predict sales price

## **3 Classification**

AI techniques to predict house quality



# BUSINESS CASE

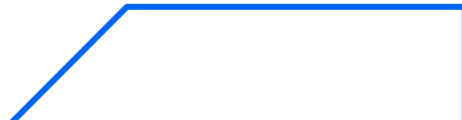
Housing market in Ames, USA



# BUSINESS CASE

Property assets are a **safe investment** method that people have trusted for years. Owning and selling a property is a common mean of **accumulating wealth**.

The main features considered when investing in the real estate market are the property **sales price** and the **quality**. They are strongly correlated.



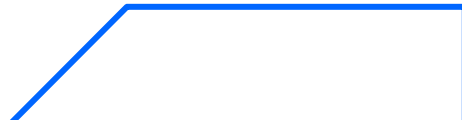




# BUSINESS CASE

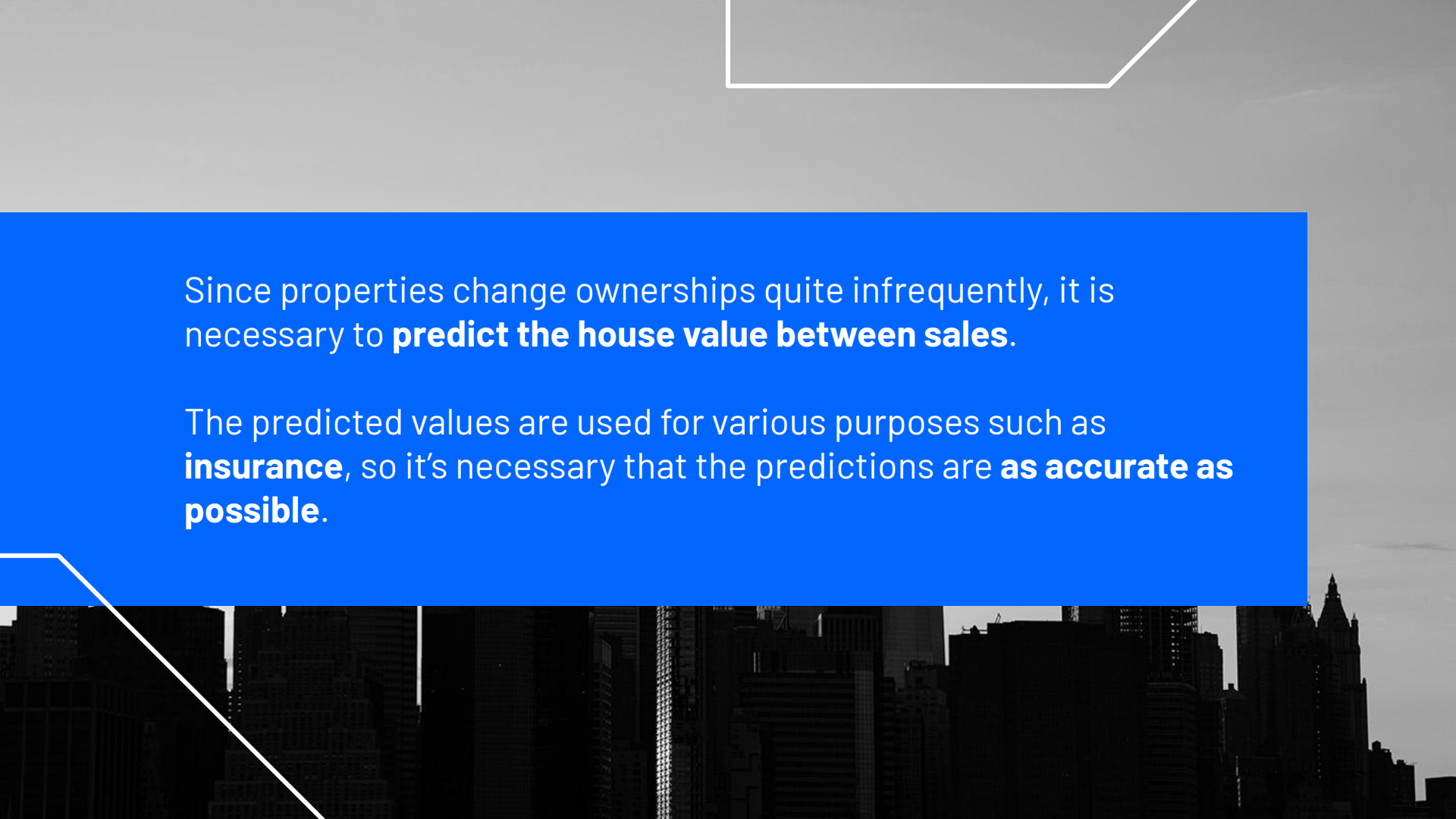
The **dataset** used contains various variables describing almost every aspect of residential homes in Ames, Iowa.

Each variable represents a house feature such as:

- Overall quality and condition
  - Square feet area
  - Sales price...
- 

# REGRESSION

AI algorithms to predict houses sales price



Since properties change ownerships quite infrequently, it is necessary to **predict the house value between sales**.

The predicted values are used for various purposes such as **insurance**, so it's necessary that the predictions are **as accurate as possible**.



# REGRESSION

Various kinds of **regression** algorithms have been used to predict sales price:

- MLR
- Ridge
- Lasso
- Random Forest
- Gradient Boosting

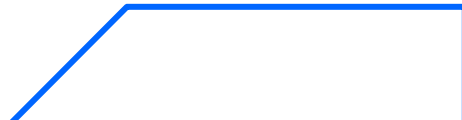




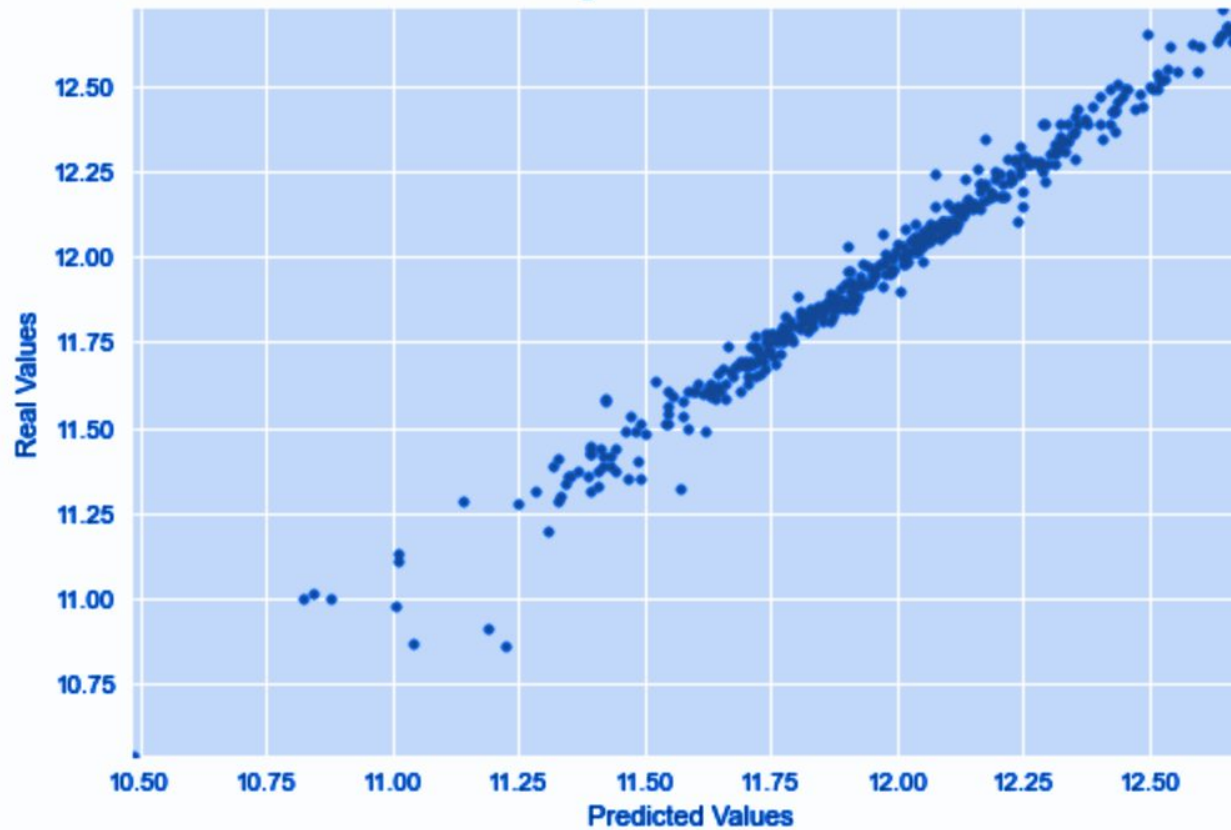
# REGRESSION

The best performing model has been the **Gradient Boosting**, followed by the Random Forest regressor.

Gradient Boosting test metrics:

- R2 score: 0.98
  - MAE: 0.03
  - RRMSE: 0.05
- 

Gradient Boosting - Predicted vs Real Sale Prices





# REGRESSION

## MLR

R2 Score Test: 0.956

## Ridge

R2 Score Test: 0.958

## Lasso

R2 Score Test: 0.960

## Random Forest

R2 Score Test: 0.973

# CLASSIFICATION

AI algorithms to predict houses quality



There are more ways of **classifying property assets** for investment and legal purposes, but one of the most common ones depends on the **quality**.

Houses are usually divided into:

- **Simple** quality (below average)
- **Standard** quality
- **Superior** quality (above average)

This information is vital to know if we are making a good financial transaction.





# CLASSIFICATION

In the original dataset, the properties quality was rated in an ordinal scale from 1 (worst) to 10 (best quality).

We converted this scale to use the **standard 3 types classification method** (simple / standard / superior) used in the real estate market.



# CLASSIFICATION

## Simple

**140 samples**  
(1-4)

avg price 103.907

## Standard

**1078 samples**  
(5- 7)

avg price 163.618

## Superior

**165 samples**  
(8-10)

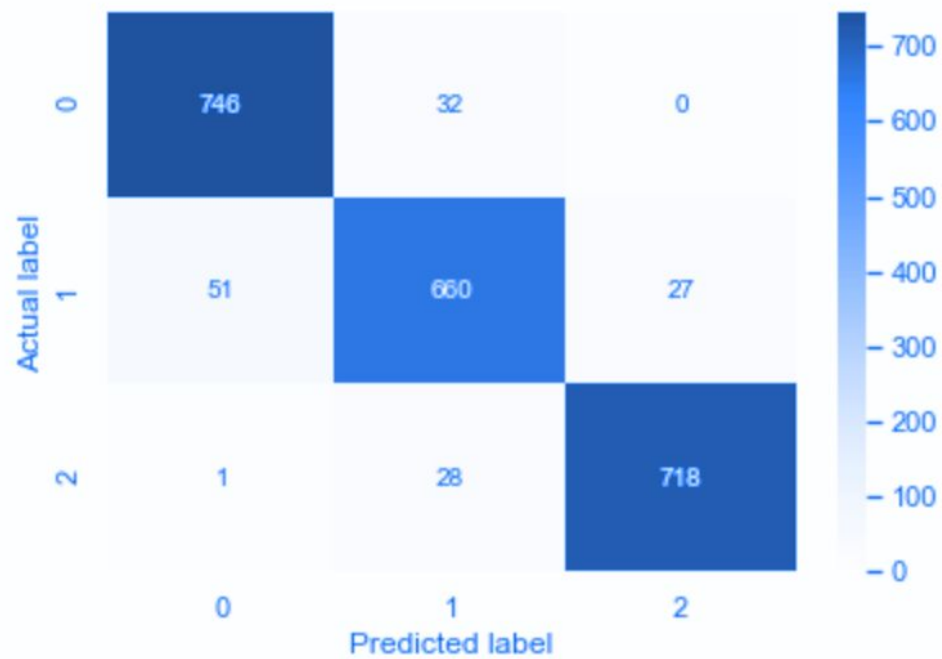
avg price 263.340

To make predictions an oversampling was needed as the data is quite unbalanced.

# CLASSIFICATION

	Accuracy	Precision	Recall	F1 Score
<b>RF</b>	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>
<b>SVM</b>	0.93	0.93	0.93	0.93
<b>Logistic</b>	0.92	0.92	0.92	0.92
<b>KNN</b>	0.79	0.79	0.80	0.78

Confusion matrix



With both Gradient Boosting regression and Random Forest classification models we can predict the sale price and the quality of a property with a high percentage of success.





A low-angle, black and white photograph of several skyscrapers reaching towards a grey sky. The perspective creates a sense of height and scale. A solid blue horizontal band is superimposed across the middle of the image, containing the text.

# THANK YOU

*Source: Ames Housing dataset, Dean De Cock, 2011*