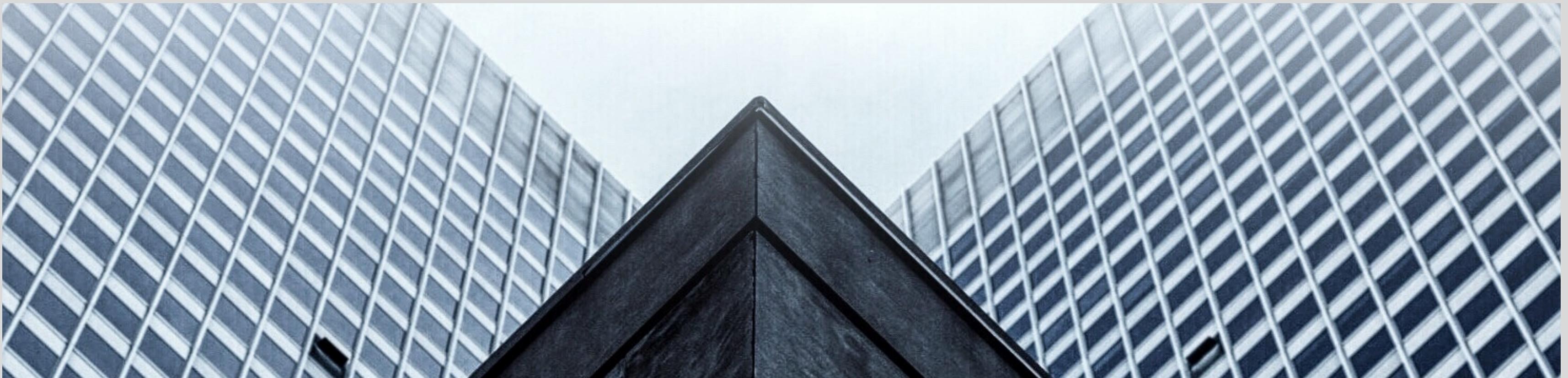


01

Using KNN with Scikit-Learn to  
classify S&P500 companies'  
industry

# Economic Sector Classification

Marcos Ribeiro | December 2019



02

# QUESTION

Can we use historical financial indicators of S&P 500 companies to classify their economic sector?

## The stock index has 11 different sectors

The Global Industry Classification Standard (GICS) is an industry taxonomy developed in 1999 by MSCI and Standard & Poor's (S&P) for use by the global financial community.

## Its is a global third party capital markets standard

GICS is used as a basis for S&P and MSCI financial market indexes in which each company is assigned to a sub-industry, and to an industry, industry group, and sector, by its principal business activity.

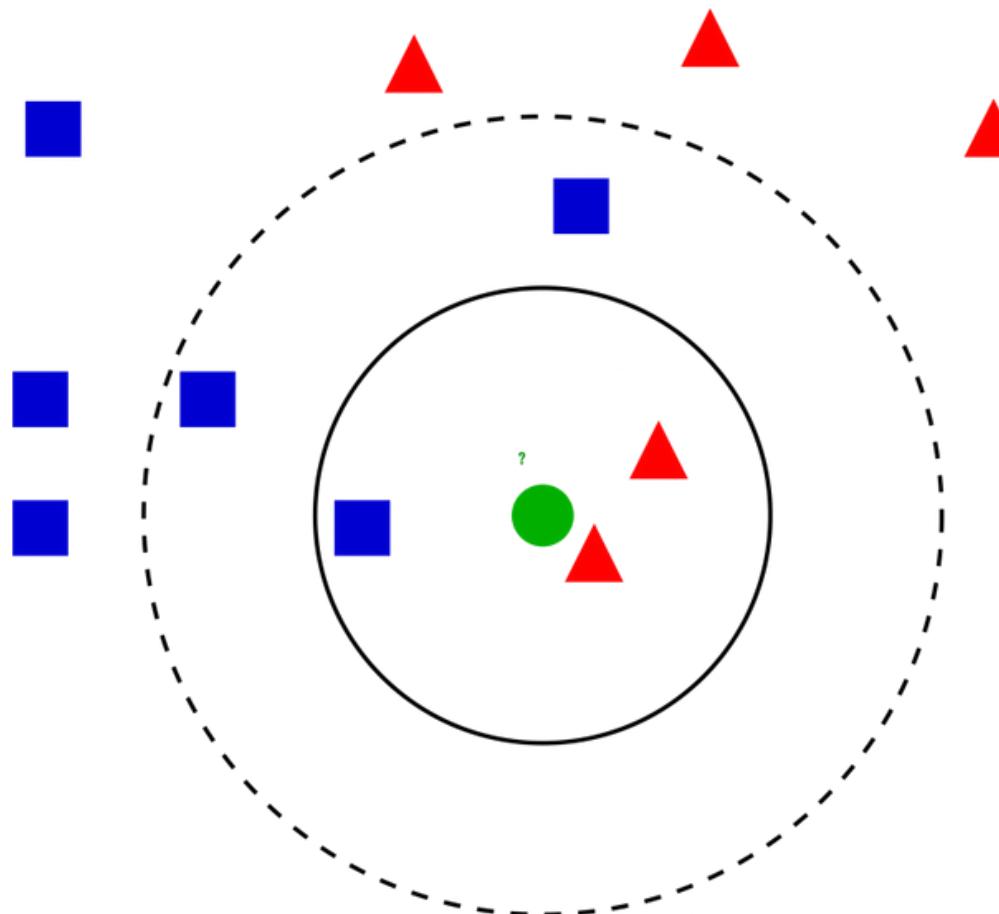
## Is financial performance a good predictor?

On this project we assess if certain industries have common financial ratios that can be used to assign a sector.

# 03 CONCEPT

How the ML learning KNN was structured

Label	Sector
0	Industrials
1	Heath Care
2	Information Technology
3	Consumer Discretionary
4	Consumer Staples
5	Utilities
6	Materials
7	Real Estate
8	Energy
9	Telecom



## Step 1

Create dataset with 10 years of complete historical financial ratios for 362 companies.

## Step 2

Perform model feature selection to reduce overfitting, improve accuracy and reduce training time.  
Selected 5 ratios out of 28.

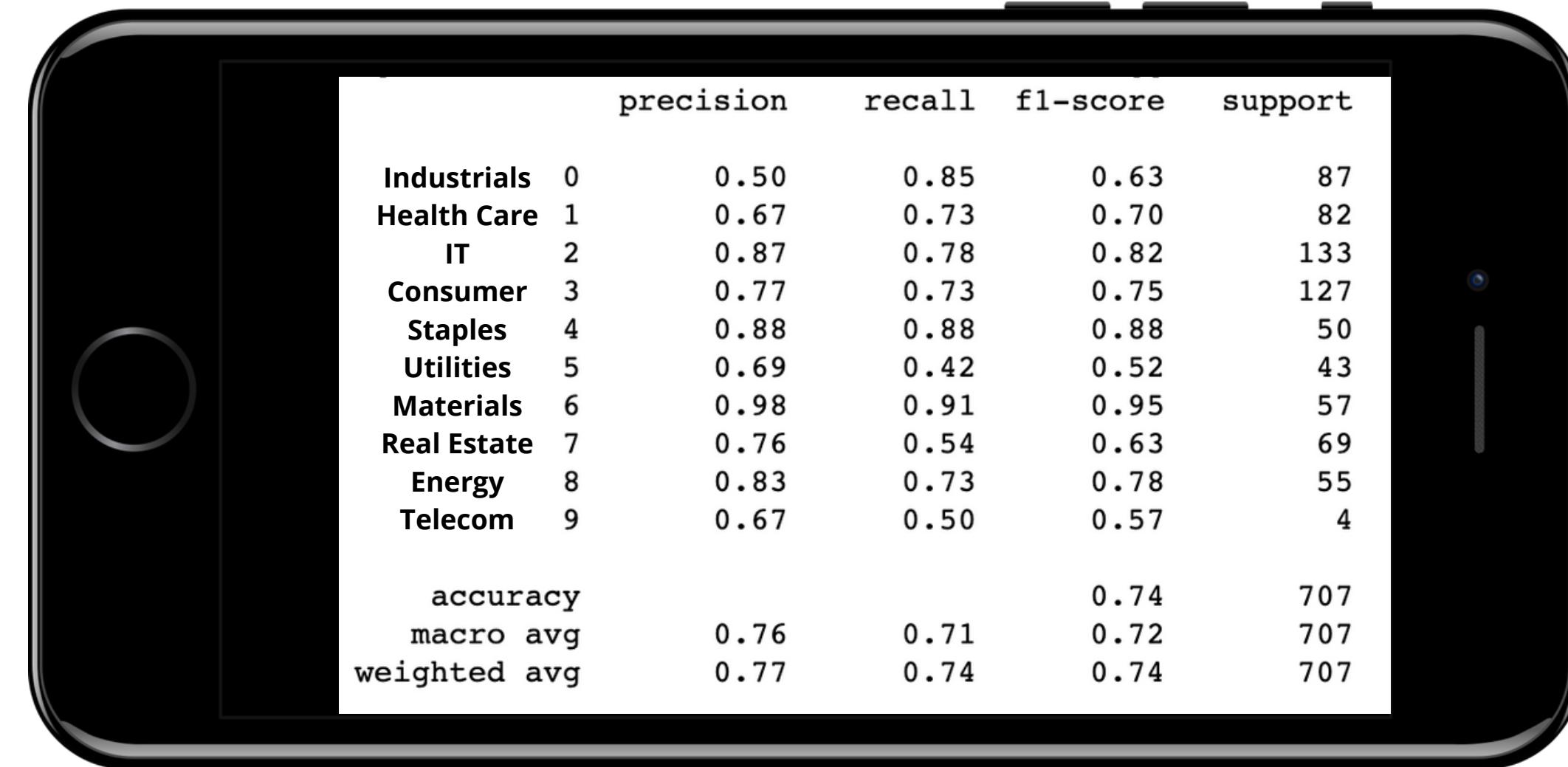
## Step 3

Confirm model's performance over unseen data with a Confusion Matrix.

# 04 Model Results

## Models' Financial Ratios

r_and_d_ratio	0.075755
asset_turn_over	0.072832
gross_margin	0.062170
ebitda_margin	0.061003
debt_assets	0.047036
current_ratio	0.043979
sg_a_ratio	0.043147
cash_ratio	0.042301
wk_turnover	0.039417
op_expense_ratio	0.036801



Precision:

"For all instances classified positive, what percent was correct?"

Recall:

"For all instances that were actually positive, what percent was classified correctly?"

F1-Score:

"Is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0."

# 05

## Future Roadmap

This model is a first step in determining which are the most effective financial ratios when assessing companies.

Other useful corporate finance applications can be derived from.

PORTFOLIO  
FRONTIERS

COMPANY  
SELECTION

CREDIT RATING  
MODELING

STOCKS  
PERFORMANCE

**THANK YOU**