

**FUJIFUSION**



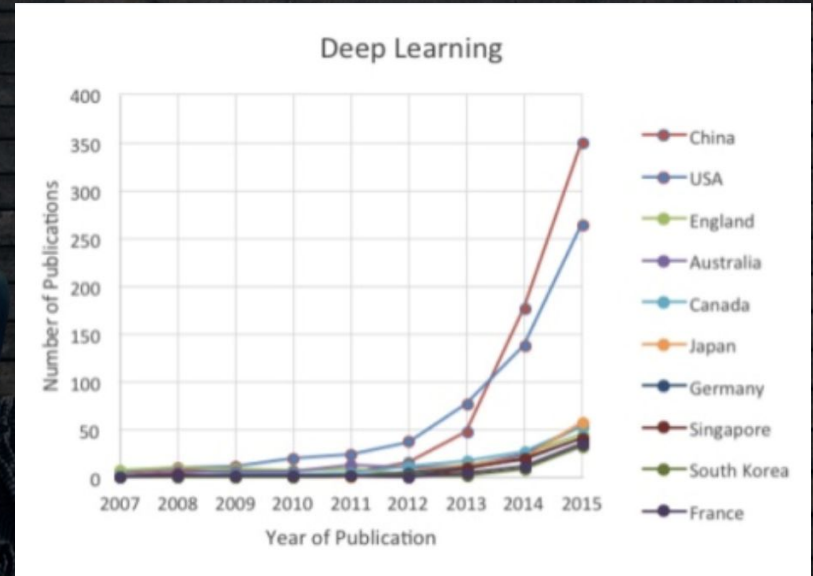
# Problem

Scholars and regulators generally agree that credit rating agency failures were at the center of the 2007-08 global financial crisis<sup>[1]</sup>.

[1] Harvard Law School Forum on Financial Regulation, "What's (Still) Wrong with Credit Ratings" (2017)

→ Financial incentive for RBC and clients to have access to accurate corporate credit ratings

→ Abundance of data and new forms of analysis (machine learning) can reduce reliance on external agencies





# Solution

Train a **machine learning model** to automate the prediction of corporate credit ratings.

Compare vendor ratings with predicted ratings to identify discrepancies.

Present this information in a **cross-platform application** for RBC's traders and clients.

# Data

date	Tickers	company	L_term_rating	L_term_numrating	Debt/EBITDA	FFO/Total debt	EBITDA/Interest	FFO/Inter
2009/12/31	XOM	EXXON MOBIL CORP	AAA	11.0	-0.0329	0.07837633870653178	297.47222222222223	256.75
2010/12/31	XOM	EXXON MOBIL CORP	AAA	11.0	0.131	0.09023499809785691	141.48181818181817	122.90909
2011/12/31	XOM	EXXON MOBIL CORP	AAA	11.0	0.0627	0.07912135660098175	206.54666666666665	179.66666
2012/12/31	XOM	EXXON MOBIL CORP	AAA	11.0	0.0304	0.08671785857464459	147.59459459459458	126.66666
2014/12/31	XOM	EXXON MOBIL CORP	AAA	11.0	0.4769	0.06547566036727642	146.47058823529412	162.17647
2009/12/31	CVX	CHEVRON CORP	AA	9.0	0.064	0.08640022203719123	99.86075949367088	78.810126
2010/03/31	CVX	CHEVRON CORP	AA	9.0	0.0948	0.10474753018660812	122.55263157894737	100.44736
2010/06/30	CVX	CHEVRON CORP	AA	9.0	0.029	0.11965572738086908	129.4935064935065	111.03896

**20 features** recorded for **27 companies**  
at multiple points in time for a total of  
**524 samples**

# Data Analysis

1| Use of the **k-nearest-neighbors** decision tree model run with a 70-30 data partition.

2| Use of **k-means** and **MeanShift** unsupervised learning algorithms to attempt to cluster companies into rating categories (with true ratings discarded)

# Results

## Prediction Accuracy

<i>Model</i>	<i>Accuracy Metric</i>	<i>Value</i>
Linear Regression	R <sup>2</sup>	0.598
K-nearest-neighbors	Jaccard similarity	0.707
<b>KNN w/ Added Features</b>	<b>Jaccard similarity</b>	<b>0.778</b>

## Clustering Results

<i>Model</i>	<i>Average cluster rating standard deviation</i>	<i>Number of clusters identified</i>
K-means	0.89	12
MeanShift	1.03	7

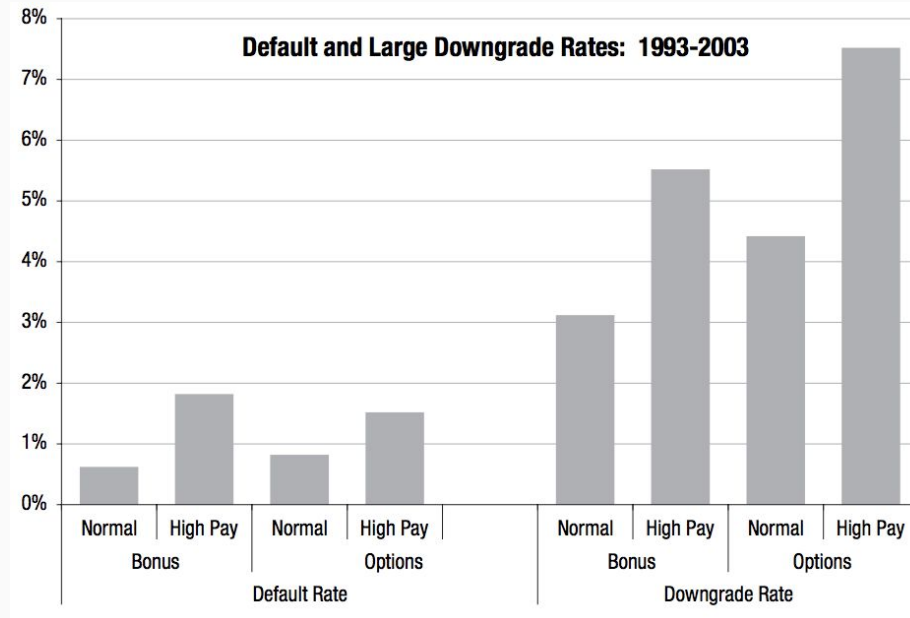
# Proposed Features

Adding the following features to the initial dataset provided by RBC increased prediction accuracy by **7%**

1. CEO Compensation
2. Median Worker Salary
3. Pay Ratio of CEO:Worker



# Why these features?



[2] Moody's Global Credit Research: "CEO Compensation and Credit Risk" (2005)

Demo!

# Next Steps to Drive Impact

## Step 1

Gather more training data. This dataset was fairly small.



## Step 2

Work with traders to do domain-specific feature engineering and selection.



## Step 3

Pilot application by deploying internally with RBC's sales & trading team.



# Team



Stephen Li

UBC  
Computer Science



Luann Jung

MIT  
Computer Science (intended)



Shekhar Kumar

UofT  
Systems Engineering



A data-driven application for  
predicting corporate credit ratings