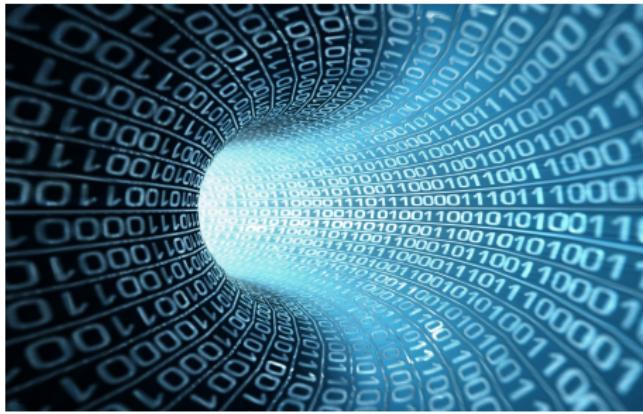


AI and Data Science Consulting

John H. Muller

November 12, 2025



Outline

1 Introduction

- About Me
- Skills, Experience, and Tools

2 Business Problems and Solutions

- Classification
- Forecasting and Regression
- Recommendations
- Generative AI
 - Generating Text
 - Summarizing Text

3 Conclusion and next steps

4 Conclusion and next steps

Introduction

About Me



- Computer Scientist (PhD), data scientist and finance quant,
- Over 15 years of experience in industry
 - Financial Services
 - Quantitative Investing
 - Retail
 - Image and video compression
- Passion for making my work matter.

Introduction

Skills, Experience and Tools

- Computer Science: algorithms and data structures
- Data Science
 - Supervised learning, clustering, regression, tree based methods
 - Neural Networks, Deep Learning and Generative AI
- Coding: Python and the python ecosystem
 - pandas, numpy, scikit-learn, pytorch ...
- SQL
- LangChain, Langgraph, CrewAI, Hugging Face

Business Problems and Solutions

Classification

What: Recognize "depression related" posts on Reddit.

How: Use a NN to *embed* posts and classify based on distance to prototypical embeddings. Similar to *Nearest Neighbor* classification.

	clean_text	is_depression
4636	mtsiaklides aw i wish i could i can t really s...	0
3492	heartbreaking to see kid taking their life out...	1
2332	maybe i should have been locked away for the r...	1
	clean_text	is_depression
6594	http twitpic com y z see where we ve been move...	0
1515	For about a week now i ve been experiencing ex...	1
209	someone pls tell me how to get over this i m c...	1

Figure 2.1: Example text from training and validation sets

Business Problems and Solutions

Classification

Examples showing embeddings in 2-D

Layer (type)	Output Shape	Param #
embedding_6 (Embedding)	(None, 100, 10)	100000
flatten_6 (Flatten)	(None, 1000)	0
dense_12 (Dense)	(None, 4)	4004
dense_13 (Dense)	(None, 1)	5

Figure 2.2: Four layer NN Model

Business Problems and Solutions

Classification

Layer (type)	Output Shape	Param #
embedding_6 (Embedding)	(None, 100, 10)	100000
flatten_6 (Flatten)	(None, 1000)	0
dense_12 (Dense)	(None, 4)	4004
dense_13 (Dense)	(None, 1)	5

Figure 2.3: Four layer NN Model

Business Problems and Solutions

Classification

How well does it work? Is it **useful**

Note that the bar for **useful?** is dependent on the application. E.g. nuclear power vs customer recommendations. give accuracy or error numbers.

Business Problems and Solutions

Classification

What: Classify images, e.g. recognize the "type" of image.

daisy



daisy



sunflowers



sunflowers



Business Problems and Solutions

Classification

How: multi-layer convolutional NN.

```
model = tf.keras.Sequential([
    tf.keras.layers.Rescaling(1./255, input_shape=(new_wd, new_wd, 3)),
    tf.keras.layers.Conv2D(16, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(16, 4, activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    #tf.keras.layers.Dropout(0.1), not active
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(1, activation="sigmoid"),
])
```

Figure 2.5: Keras model

Business Problems and Solutions

Classification

How well does it work? Is it **useful**?

Give example of accuracy. Also, maybe show some example misclassification

Business Problems and Solutions

Forecasting and Regression

What: forecast the monthly **change** in non-farm payroll.

Data Source: BLS via St. Louis Fed site FRED.

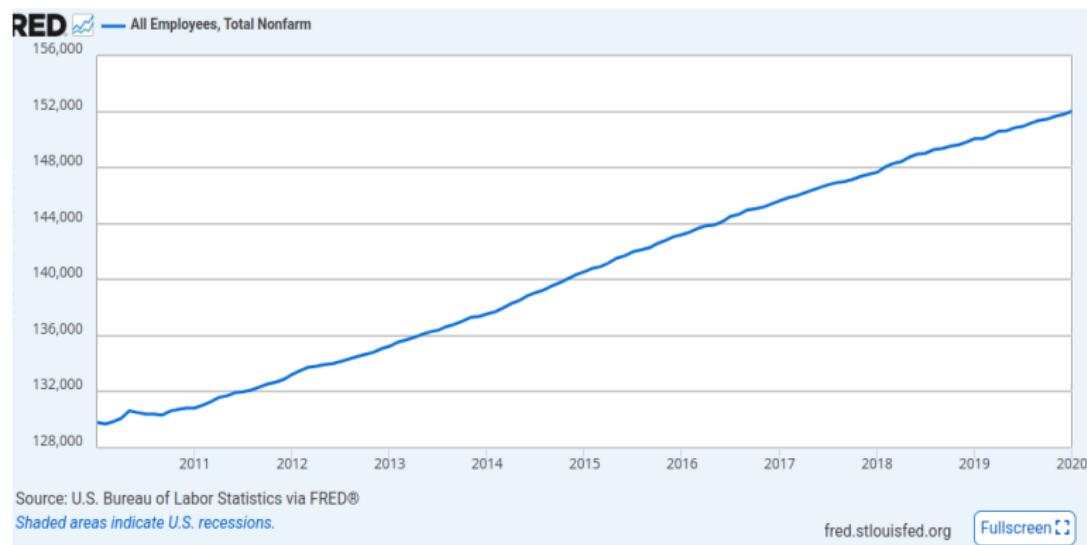


Figure 2.6: NonFarm Payroll level, 2010 - 2020

Business Problems and Solutions

Forecasting and Regression

Not always a steady increase.

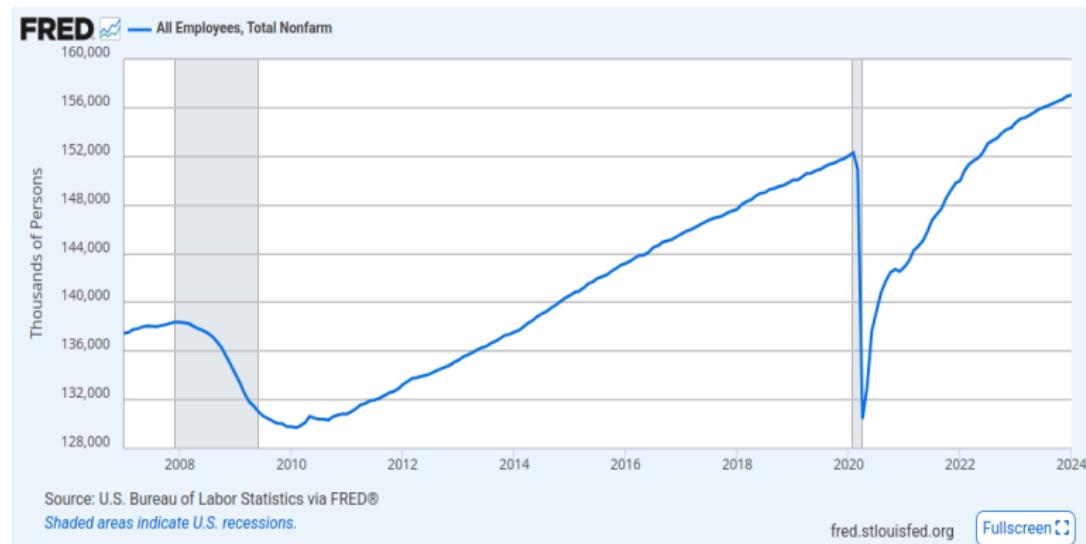


Figure 2.7: NonFarm Payroll level, 2008 - 2023

Business Problems and Solutions

Forecasting and Regression

Changes in the level look less predictable.

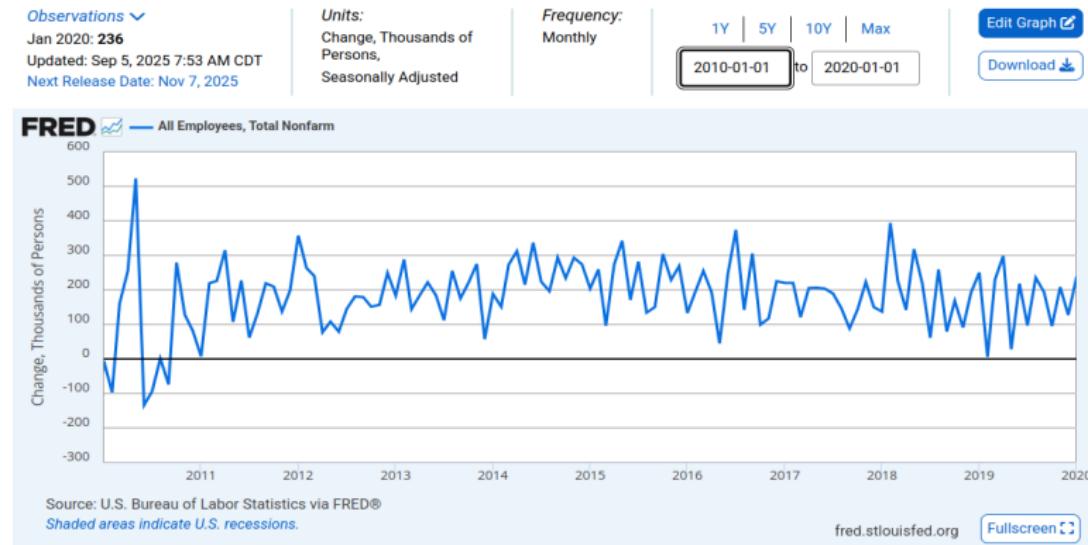


Figure 2.8: NonFarm Payroll, changes

Business Problems and Solutions

Forecasting and Regression

What: forecast the next month's change in non-farm payroll.

How: regress the monthly change on

- One or more lagged values.
- Unemployment rate.
- Civilian Labor force.
- Other data that might be helpful:
 - GDP (quarterly)
 - unemployment claims (weekly)
 - ADP payroll changes (monthly)
 - Google trend values for "job", "layoff", ... (hourly)

Business Problems and Solutions

Forecasting and Regression

How well does it work? Is it **useful?**

Useful if it can reliably predict whether the number will beat expectations.
show plot of actual vs. predicted

Business Problems and Solutions

Recommendations

What: predict a viewer's rating for a movie.

How: Embedding NN for both viewer and the movie.

The similarity between viewer and rating determines the predicted rating.

Business Problems and Solutions

Recommendations

Simple explanation of the model.

Back to the plot explaining embedding.

Business Problems and Solutions

Generative AI

What: Given a minimal set of *facts*, generate a narrative or story about the facts.

Business Problems and Solutions

Generative AI

What: Given a minimal set of *facts*, generate a narrative or story about the facts.

Conclusion and next steps

Let me help you solve your problems.
Next steps?

Conclusion and next steps

Thank you!

John H. Muller

jmuller.ics88@gtalumni.org

+1 (617) 669-2204