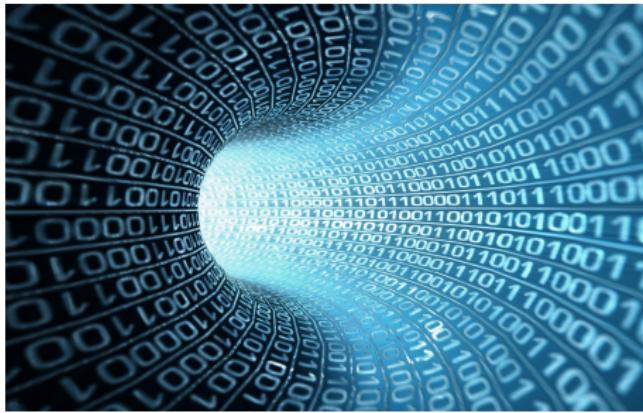


AI and Data Science Consulting

John H. Muller

November 13, 2025



Outline

1 Introduction

- About Me
- Skills, Experience, and Tools

2 Business Problems and Solutions

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

3 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: Compare 3 methods of classifying on a medical dataset.

For details on the dataset see Kaggle

<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

Business Problems and Solutions

Classification

Variables

- ① id: unique identifier
- ② gender: "Male", "Female" or "Other"
- ③ age: age of the patient
- ④ hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
- ⑤ heart_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
- ⑥ ever_married: "No" or "Yes"
- ⑦ work_type: "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
- ⑧ Residence_type: "Rural" or "Urban"
- ⑨ avg_glucose_level: average glucose level in blood
- ⑩ bmi: body mass index
- ⑪ smoking_status: "formerly smoked", "never smoked", "smokes" or "Unknown"*
- ⑫ stroke: 1 if the patient had a stroke or 0 if not

Business Problems and Solutions

Classification

- logistic regression
- gradient boosting
- random forests

Business Problems and Solutions

Classification

Accuracy and loss on train, validation and test datasets. Loss is binary cross-entropy loss.

acc	loss	ds
0.950610	0.156825	train
0.948946	0.173313	val
0.953815	0.135607	test

Figure 2.1: Logistic Regression

acc	loss	ds
0.964761	0.094885	train
0.958935	0.119844	val
0.959839	0.104992	test

Figure 2.2: RandomForest

Business Problems and Solutions

Classification

acc	loss	ds
0.950888	0.151802	train
0.948946	0.168683	val
0.953815	0.136938	test

Figure 2.3: Gradient Boosting

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.

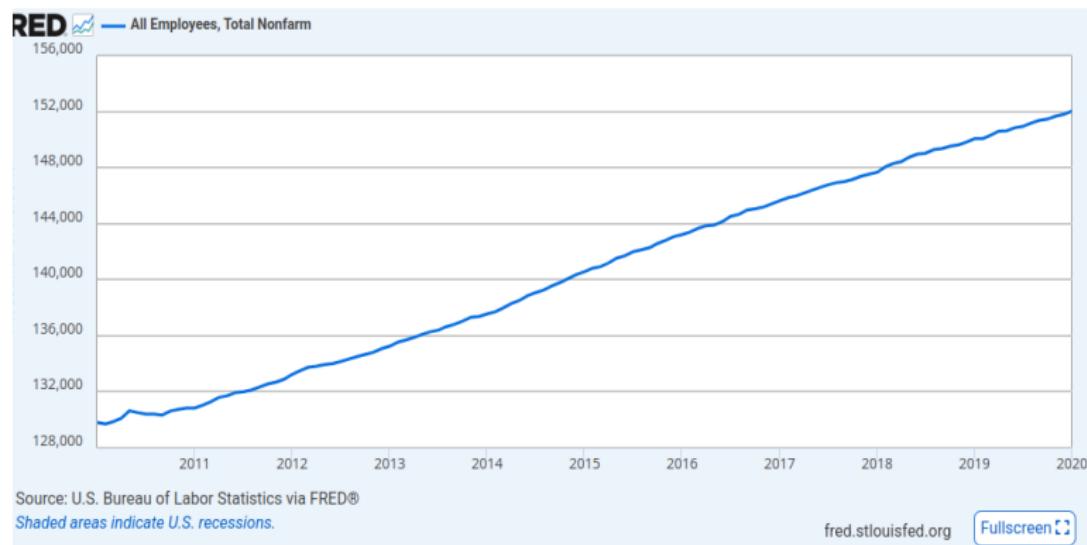


Figure 2.4: NonFarm Payroll level, 2010 - 2020

Business Problems and Solutions

Forecasting and Regression

Not always a steady increase.

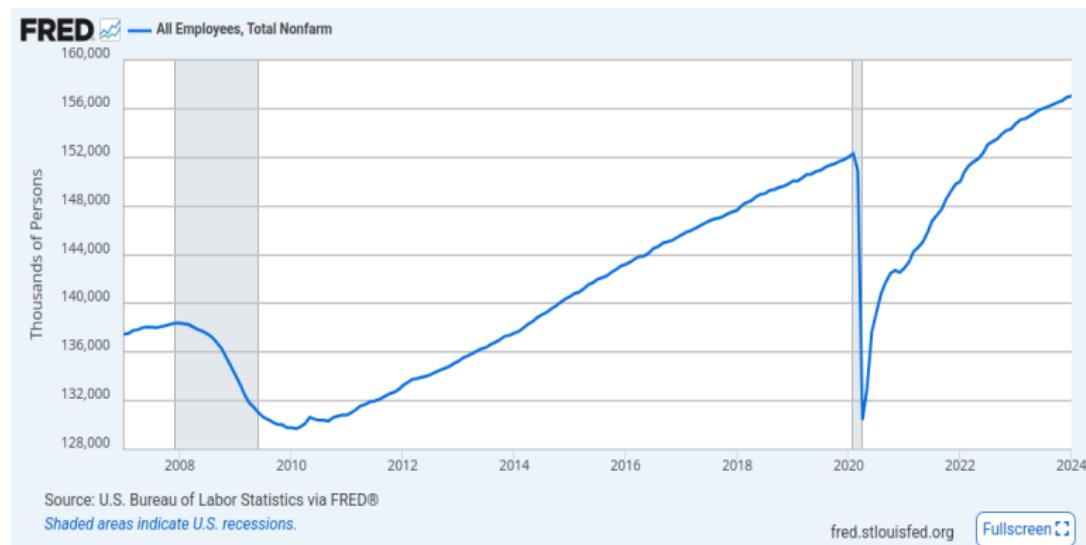


Figure 2.5: NonFarm Payroll level, 2008 - 2023

Business Problems and Solutions

Forecasting and Regression

Changes in the level look less predictable.

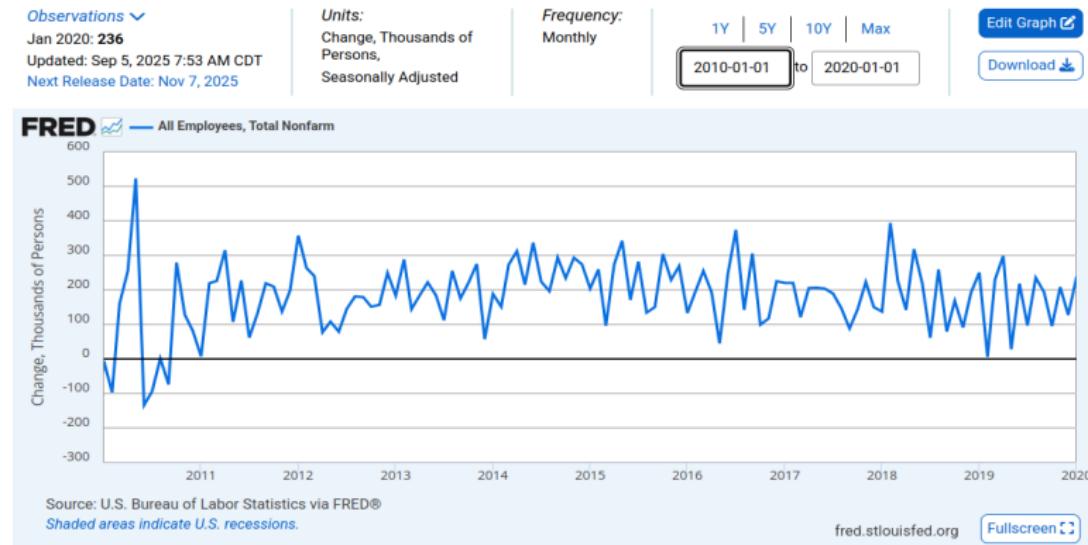


Figure 2.6: NonFarm Payroll, changes

Business Problems and Solutions

Forecasting and Regression

What: forecast the next month's change in non-farm payroll.
How: regress the monthly change on

- ① One or more lagged values.
- ② Unemployment rate.
- ③ Civilian Labor force.
- ④ Other data that might be helpful:
 - GDP (quarterly)
 - unemployment claims (weekly)
 - ADP payroll changes (monthly)
 - Google trend values for "job", "layoff", ... (hourly)

I will start with the first 3 above.

Business Problems and Solutions

Forecasting and Regression

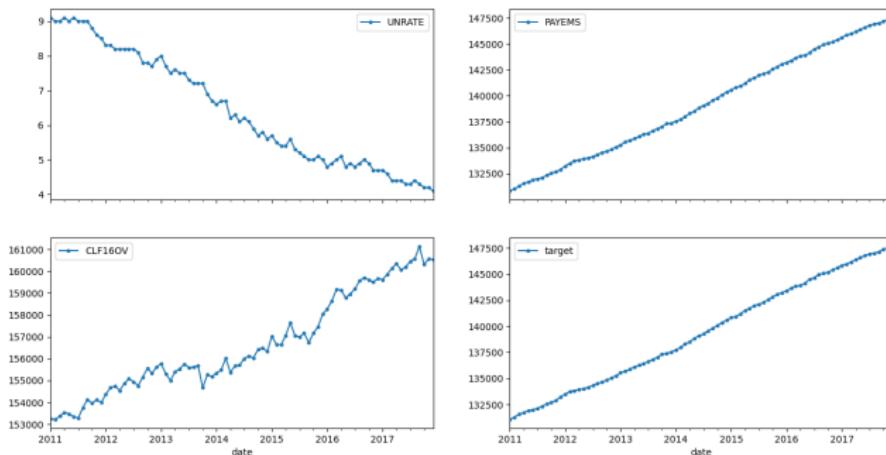


Figure 2.7: target and input variables, level

Business Problems and Solutions

Forecasting and Regression

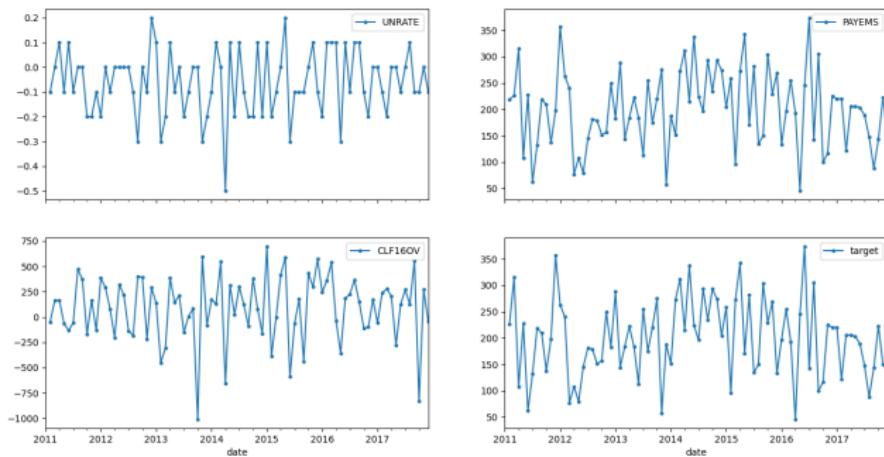


Figure 2.8: target and input variables, changes

Business Problems and Solutions

Forecasting and Regression

How well does it work? Is it **useful**? See the true vs. predicted values below. Traders often react to whether the number beat the expected number. We can use 200 as a proxy for expectation, so we accurately predicted above or below about half the months.

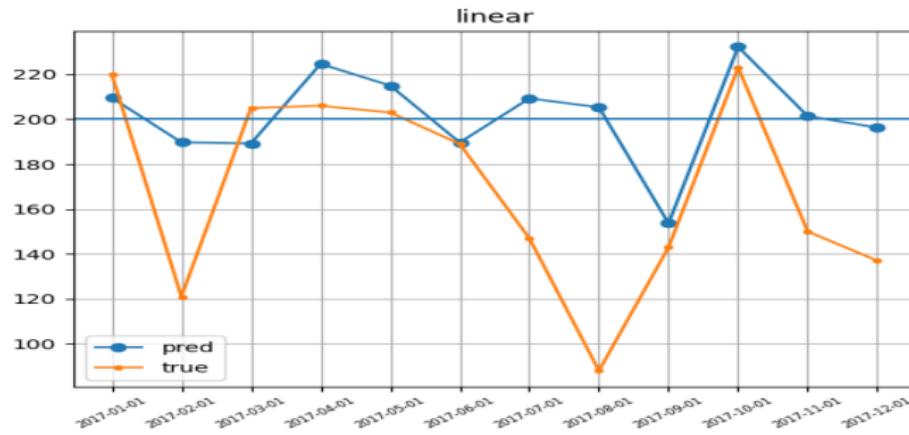


Figure 2.9: results using a linear model

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

Pytorch model

```
1 class MF(nn.Module):
2     def __init__(self, n_users, n_movies, emb_size=100):
3         super(MF, self).__init__()
4         self.n_users = n_users
5         self.n_movies = n_movies
6         self.user_emb = nn.Embedding(n_users, emb_size)
7         self.movie_emb = nn.Embedding(n_movies, emb_size)
8
9         # initializing the matrices with a positive number
10        # supposed to help generally will yield better results
11        self.user_emb.weight.data.uniform_(0, 0.5)
12        self.movie_emb.weight.data.uniform_(0, 0.5)
13
14    def forward(self, users, movies):
15        m = self.movie_emb(movies)
16        u = self.user_emb(users)
17        return (u * m).sum(1) # taking the dot product
```

Business Problems and Solutions

Recommendations

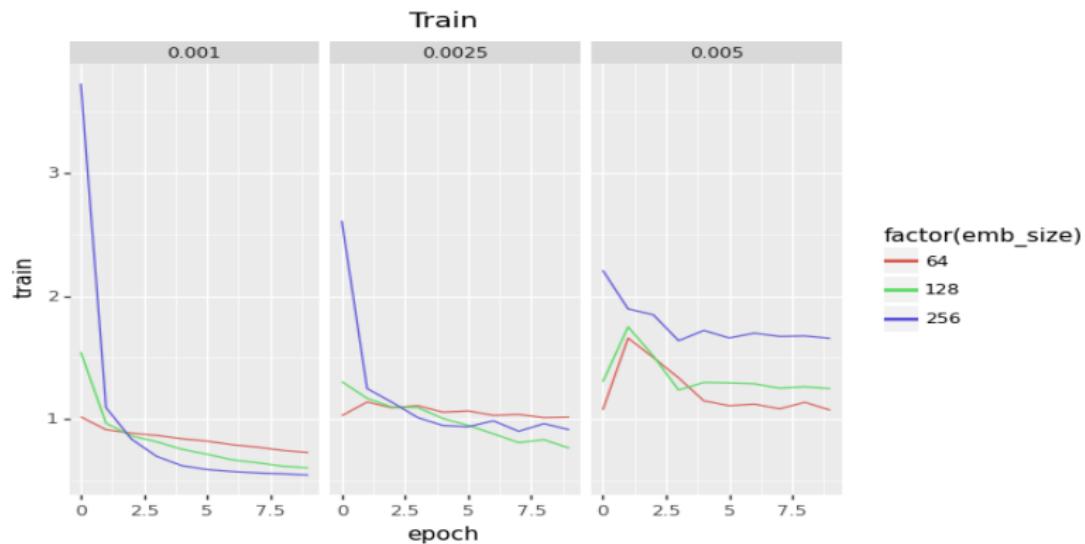
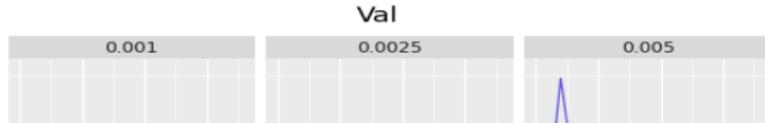


Figure 2.10: Interface to gather 5 basic facts



Business Problems and Solutions

Recommendations

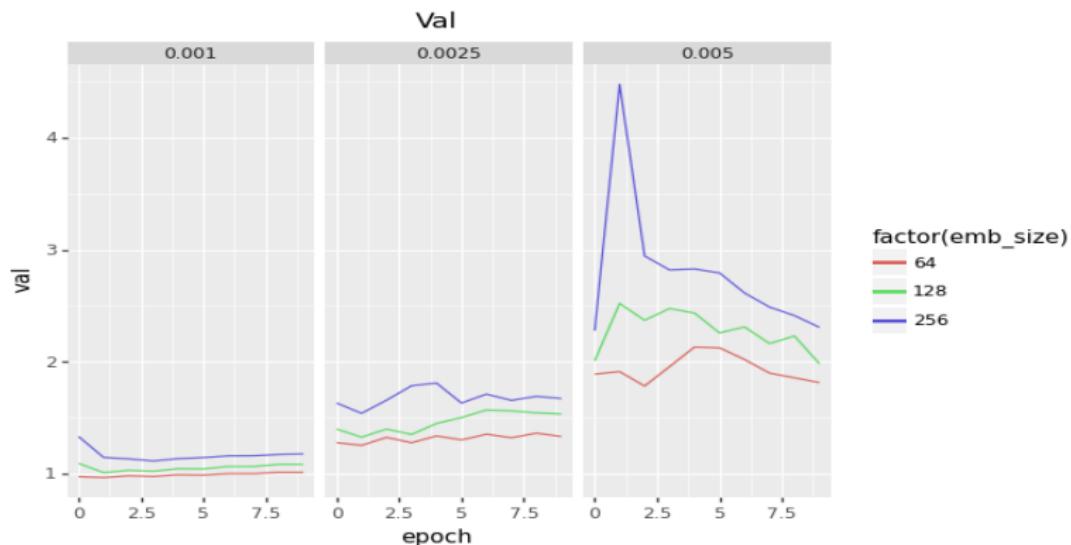


Figure 2.12: Interface to gather 5 basic facts

Test error was 1.23 with 64 dimensional embeddings

Business Problems and Solutions

Generative AI

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

Business Problems and Solutions

Generative AI

The interface consists of several input fields:

- airline:** A radio button group with "Delta" selected and "American" as an option.
- travel_date:** A text input field containing "May 2".
- tone:** A radio button group with "polite" selected and four other options: "polite but firm", "annoyed", "livid", and "litigious".
- issues:** A text input field containing "lost luggage" with a clear button ("x") to its right.
- compensation:** A radio button group with "cash" selected and two other options: "reward miles" and "free flight".

At the bottom are two buttons: "Clear" (gray) and "Submit" (orange).

Figure 2.13: Interface to gather 5 basic facts

Business Problems and Solutions

Generative AI

Dear Delta Customer Service Team,

I hope this message finds you well. My name is John Smith, and I am writing to address some issues I encountered during my recent travel experience with Delta Airlines. My Frequent Flyer number is 2194048289.

On May 2, I was scheduled to fly from Montreal to Atlanta. Unfortunately, upon arrival in Atlanta, I discovered that my luggage had not arrived with me. This was quite distressing, as I was left without my belongings for an extended period. I understand that lost luggage can happen, but the lack of communication and updates regarding the status of my bag added to the inconvenience.

Further complicating my travel experience, the delay in receiving my luggage impacted my plans significantly. I had to incur unexpected expenses to replace some essentials during that time, which was both frustrating and financially burdensome.

I appreciate the efforts Delta Airlines makes to ensure customer satisfaction, and I would like to kindly request compensation for the challenges I faced during this journey. Cash compensation would greatly assist in alleviating the financial impact caused by the incident and would reaffirm my trust in Delta Airlines as a valued airline.

Thank you for taking the time to consider my request. I look forward to your prompt response and a resolution to this matter.



Figure 2.14: resulting "polite" letter

Business Problems and Solutions

Generative AI

Dear Customer Service,

I am writing to you a very frustrated and dissatisfied frequent flyer (Frequent Flyer Number: 2194048289) regarding the appalling experience I had on May 2 when I was scheduled to fly from Montreal to Atlanta on your airline.

To say my experience was disappointing would be an understatement. Upon arrival, I was not only met with the usual travel chaos, but also the absolute nightmare of dealing with lost luggage. Not only was my bag essential for my trip, but it also contained personal items that I cannot simply replace. This situation is unacceptable and reflects poorly on the level of service Delta Airlines claims to provide.

I navigated the airport and your customer service lines only to receive little more than empty promises and a lack of urgency in resolving the issue. I expected better from a reputable airline such as yours. Losing luggage is a huge inconvenience, and being left in the dark with no real solutions further compounded my frustration.

I am compelled to request an appropriate cash compensation for the distress this ordeal caused me. It is the least Delta Airlines could do to restore some faith in your customer service.

I look forward to your prompt response and a satisfactory resolution to this matter.

Sincerely,
John Smith

Figure 2.15: resulting "livid" letter

Business Problems and Solutions

Generative AI

What: Condense a lengthy passage of several pages down to several paragraphs. Click the link below for a demo.

▶ Link

Please note a few things:

- You will probably have to restart the app
- Be patient, I am using limited processing

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