Predicting Depression Prevalence Using Deep Learning

DATA 5610 Final Project
Spring 2025
Carter Grant

Introduction – What is the Problem and Why Does it Matter?

Depression is one of the most pervasive public health challenges in the United States and globally. According to the CDC, the rate of adults who report having been diagnosed with depression has increased significantly over the last two decades. As policymakers, public health officials, and mental health professionals look to better allocate resources and improve early intervention, understanding the demographic and geographic patterns of depression prevalence is crucial.

This topic is very personal to me. I have struggled with depression since high school, and I've seen how profoundly it can impact individuals and even entire communities. Friends, classmates, and family members of mine have experienced it, and in some cases, I've seen it take lives. These experiences have shaped my understanding of how critical it is to not only talk about mental health but to actively work toward solutions that help identify and support those at risk.

Through this project, I hope to contribute to that effort by using data to find patterns in depression prevalence. By identifying which combinations of demographic factors (such as income level, education, gender, race, and geographic location) are associated with the highest rates of depression, we can better inform the people who shape policy. My goal is to help ensure that mental health resources are targeted where they are needed most, and that vulnerable groups are not overlooked.

This project addresses the question: Can we accurately predict depression prevalence using group-level demographic and geographic data with deep learning? Specifically, I aim to identify which population subgroups are most at risk and explore how those risks vary across states over time. To explore this question, I used data from the Behavioral Risk Factor Surveillance System (BRFSS), which is a large state-level survey of adult health behaviors and conditions, including mental health diagnoses. I will develop a deep learning model that can estimate depression prevalence across subgroups and serve as a tool for public health analysis.

The proposed approach uses a feedforward neural network trained on one-hot encoded demographic, state, and year features, along with sample size normalization. My key finding is that this model significantly outperforms traditional baselines (random forest, linear regression) in predicting depression prevalence values. The also model identifies interpretable subgroup trends that align with known trends and patterns in mental health.

This paper documents my process from data selection to model evaluation and concludes with next steps for enhancing prediction power and real-world application. I have learned to better understand depression, how to counter it, and how to use neural networks better through the creation of this project, I hope you can do the same.

Approach – How Did You Address the Question?

Data:

For this project, I used data from the **Behavioral Risk Factor Surveillance System (BRFSS)**, which is a nationally representative health-related telephone survey conducted annually by the Centers for Disease Control and Prevention (CDC). Covering all 50 states, the District of Columbia, and several U.S. territories, the BRFSS is one of the most comprehensive datasets for understanding chronic health conditions and behavioral risk factors in the United States. The dataset I worked with spans from 2011 to the present and includes responses to a wide range of public health indicators, including mental health conditions such as depression.

The specific variable of interest was derived from the question: “Has a doctor or other healthcare provider ever told you that you have a form of depression?” I filtered the data to include only the responses where individuals answered "Yes" to this question. Each record represents a unique combination of year, state, and demographic groups where demographic groups are broken down by attributes such as sex, income, race, education level, and more. I combined the Break_Out_Category and Break_Out fields from the original dataset into a single Demographic feature to streamline this information. The Data_value field which represents the percentage of people in each group who reported a depression diagnosis served as the prediction target.

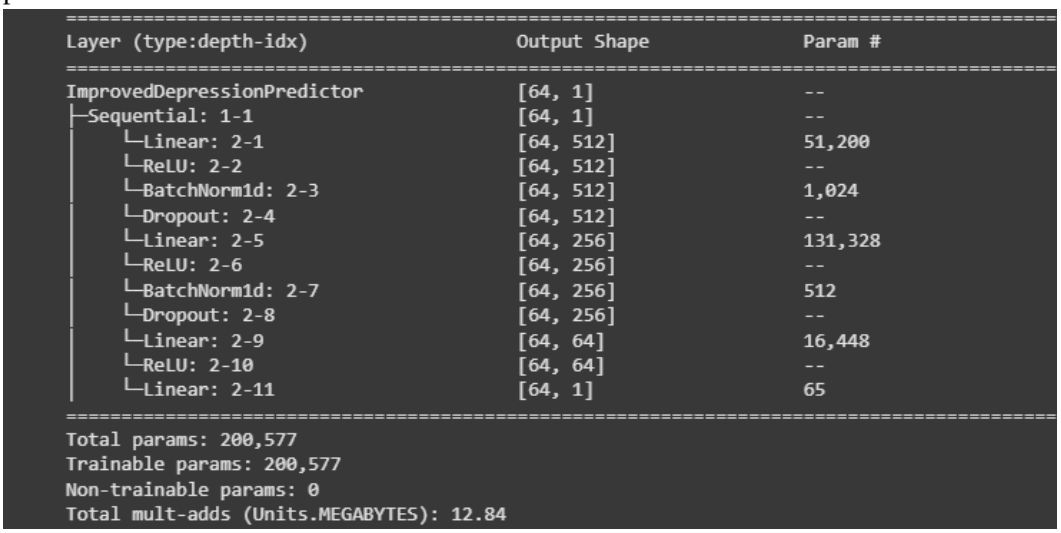
Several preprocessing steps were necessary to prepare the data for training a machine learning model. All categorical variables, including Year, Locationabbr (state), and Demographic, were one-hot encoded using scikit-learn’s OneHotEncoder. The Sample_Size field, which indicates how many individuals were surveyed within each group, was retained as a continuous feature and is standardized with StandardScaler. This ensured that the model could recognize the reliability of group-level percentages by accounting for variation in sample sizes. I then split the full dataset into training and test sets using an 80/20 ratio and converted the inputs to PyTorch tensors to feed into the neural network.

The BRFSS dataset is well-suited for this project due to its breadth, workability, and consistency. Its large sample size and demographic diversity allowed me to investigate depression prevalence across a wide range of social groups and over multiple years. However, there are some limitations to be aware of. Because the data is self-reported, it is susceptible to recall bias and underreporting, particularly in communities where mental health stigma is strong. In addition, some groups (such as high-income Asian males in rural states) appear infrequently, which may limit model generalizability for these populations. One-hot encoding, while straightforward, also fails to capture relationships between categories—such as the similarity between adjacent income levels.
Despite these limitations, the BRFSS remains a solid source for public health analysis in the U.S. Its consistency, scope, and accessibility make it an excellent dataset for exploring patterns in depression diagnosis across time, space, and social context.

Methodology:

To predict depression prevalence based on demographic and geographic information, I implemented a deep learning approach using a fully connected feedforward neural network (FNN). This architecture was chosen because of its capacity to model complex, nonlinear relationships in high-dimensional data. The input data consisted primarily of categorical variables such as year, state, and demographic groups, all of which were one-hot encoded. Additionally, I included sample size as a numerical input to help the model account for group-level confidence in prevalence estimates. Given the number of categories and potential interactions between variables, a neural network offered a flexible and powerful framework for capturing patterns that simpler models like linear regression or decision trees would likely miss.

Initially, I tested a basic model with a single hidden layer but found that it underfit the training data and failed to capture patterns in the prevalence rates. After the initial model showed signs of underfitting, I conducted a grid search over various combinations of hidden layers and dropout rates to improve performance and reduce overfitting. I compared architectures ranging from two to four layers and varied dropout values between 0.1 and 0.3. Validation loss was monitored after each epoch to select the best-performing model. Through this process, I found that a model with three hidden layers with dropout and batch normalization at each layer, offered the best trade-off between learning capacity and overfitting prevention. A summary of the model architecture, including parameter counts and layer outputs, is provided below:



Layer (type:depth-idx)	Output Shape	Param #
ImprovedDepressionPredictor	[64, 1]	--
├─Sequential: 1-1	[64, 1]	--
│ └─Linear: 2-1	[64, 512]	51,200
│ └─ReLU: 2-2	[64, 512]	--
│ └─BatchNorm1d: 2-3	[64, 512]	1,024
│ └─Dropout: 2-4	[64, 512]	--
│ └─Linear: 2-5	[64, 256]	131,328
│ └─ReLU: 2-6	[64, 256]	--
│ └─BatchNorm1d: 2-7	[64, 256]	512
│ └─Dropout: 2-8	[64, 256]	--
│ └─Linear: 2-9	[64, 64]	16,448
│ └─ReLU: 2-10	[64, 64]	--
│ └─Linear: 2-11	[64, 1]	65

Total params: 200,577
Trainable params: 200,577
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 12.84

Figure A1 – Model Architecture
Note Appendix B to see the code used to create architecture

The model was trained for 150 epochs, with mini-batches of 64 samples per update. Training and validation losses were tracked at each epoch to ensure stability and identify overfitting.

To evaluate the final model, I computed several metrics on the held-out test set: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2, and Adjusted R^2. These metrics helped me assess both accuracy and generalization. The results were also benchmarked against two baseline models which included a random forest and a linear regression model. These models were trained on the same features. While both baselines captured general trends, the neural network consistently performed better, especially in identifying patterns across underrepresented or overlapping subgroups. This confirms that the selected architecture was not only appropriate for the task but also meaningfully better than standard regression techniques.

Analysis and Results – What Did You Find?

After extensive tuning and training, the final deep learning model was evaluated on a held-out test set

using three key performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2.The model achieved an RMSE of 2.58, MAE of 1.70, and an R^2 of 0.855, which indicates strong alignment between predicted and actual group-level depression prevalence values. This suggests that the model explains approximately 86% of the variation in depression prevalence across demographic and geographic groups. To assess whether this performance justified the use of deep learning, I benchmarked the neural network against two standard models: a random forest and a linear regression model. The random forest achieved an R^2 of 0.805, while the linear regression lagged behind at 0.747. In terms of error metrics, both models had significantly higher RMSE and MAE scores compared to the neural network. This reinforces the strength of the deep learning approach in capturing subtle and nonlinear trends within this high-dimensional, sparse data.

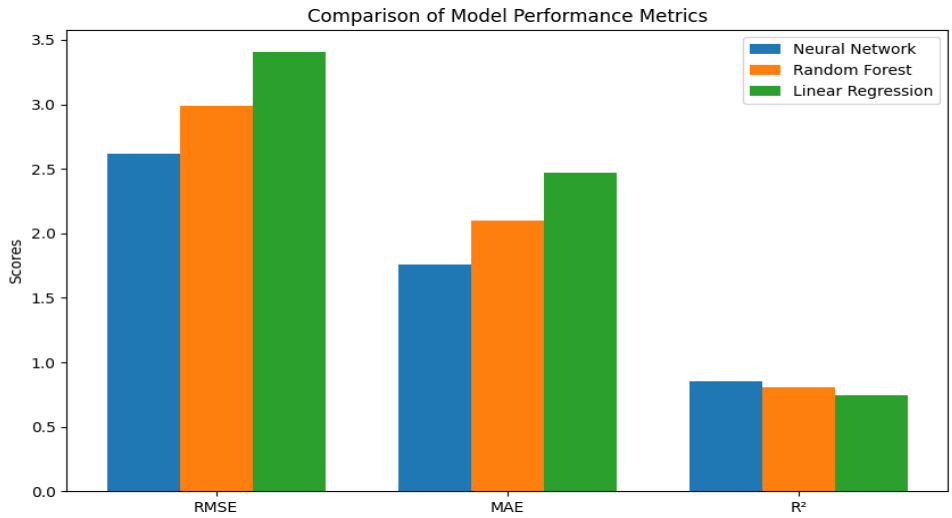


Figure A2 – Comparison of Model Performance Metrics

To further visualize how well the neural network predicted outcomes, a scatterplot of predicted vs. actual values shows that most predictions align closely with the observed prevalence percentages. Most data points cluster near the 45-degree reference line, signaling high accuracy. Notably, the model slightly underpredicts for some high-prevalence subgroups and overpredicts for groups with small sample sizes, which is a common challenge in survey-based datasets.

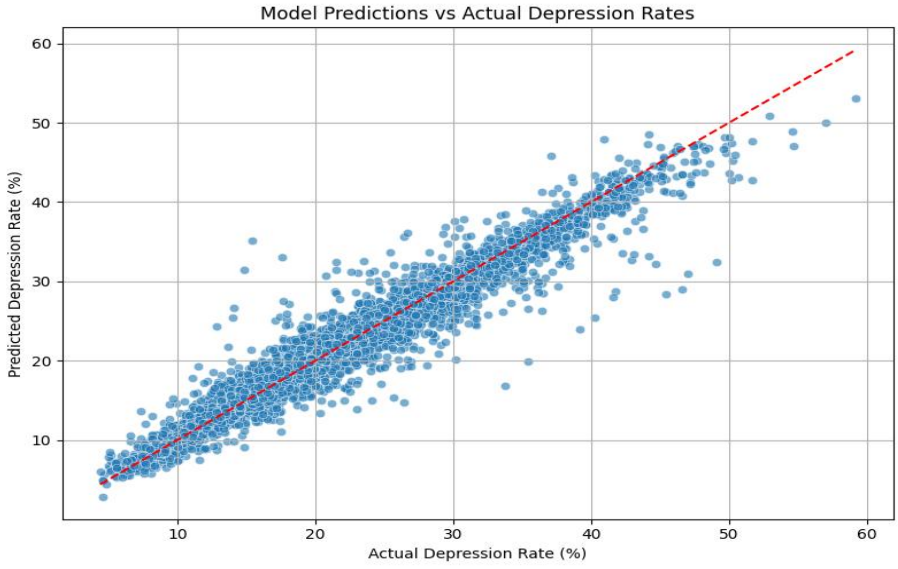


Figure A3 – Predicted vs. Actual Scatterplot

To better understand where mental health disparities appear most severe, I also decided to rank the top 10 demographic subgroups with the highest predicted prevalence. These groups consistently included low-income women, individuals from New Hampshire and Maine, and multiracial or Hispanic populations.

	Year	Locationabbr	Demographic	Predicted_Depression
13879	2022	NH	Household Income - Less than $15,000	53.137508
7783	2017	NH	Household Income - Less than $15,000	50.928761
3054	2013	NH	Race/Ethnicity - Multiracial, non-Hispanic	50.070442
13658	2022	ME	Household Income - Less than $15,000	49.278641
13001	2021	VT	Household Income - Less than $15,000	48.892654
15656	2023	WV	Household Income - Less than $15,000	48.598751
7578	2017	ME	Household Income - Less than $15,000	48.197201
14948	2023	ME	Household Income - Less than $15,000	48.168423
13882	2022	NH	Race/Ethnicity - Multiracial, non-Hispanic	47.940617
5410	2015	NH	Household Income - Less than $15,000	47.729649

Figure A4 – Top 10 At-Risk Demographic Subgroups

Beyond the strong numerical performance, several meaningful patterns emerged from model predictions that deepen my understanding of how depression affects different populations. For instance, the model clearly recognized that socioeconomic status plays a central role in mental health outcomes. Groups with low income, low education levels, and multiracial ethnicity were repeatedly ranked higher in predicted prevalence, even when controlling for gender or race. This suggests the model is sensitive not just to surface-level categories but also to deeper structural inequalities embedded in the data.
The model also performed especially well when predicting for large population groups with well-distributed sample sizes while struggling somewhat with small or highly specific subgroups (e.g., Native Hawaiian or Pacific Islander females in low-income brackets).

Another insight came from the model's apparent ability to learn from geographical variation. Although no external geographic data like rurality, policy, or climate was included, the model's predictions aligned with known mental health trends, such as elevated prevalence in the Southern U.S. and relatively lower values in the Mountain West. This indicates that even when trained solely on encoded state labels, the model captures location-based context that reflects deeper systemic and environmental influences.
The model not only achieved strong overall accuracy but also uncovered relationships that offer policy-relevant insights. It succeeded in identifying vulnerable subgroups, detecting disparities across states, and tracking national trends over time. Its performance demonstrates the value of using deep learning not just as a predictive tool, but as a means of exploring hidden structure in public health data (See Appendix A for more visuals).

Discussion, Conclusions, and Next Steps

This project set out to answer the question “Can we accurately predict depression prevalence using group-level demographic and geographic data with deep learning?” Based on the results mentioned above, the answer is yes. The neural network achieved strong predictive performance, outperforming traditional machine learning models like random forest and linear regression on all key evaluation metrics. More importantly, it identified interpretable patterns in mental health risk that reflect real-world disparities across socioeconomic, racial, gender, and geographic lines.

These findings have very important and significant implications for both research and policy. At the research level, this model demonstrates how relatively simple deep learning architectures can surface valuable insights in behavioral health domains. From a policy perspective, the model's ability to gather high-risk subgroups and highlight regional disparities offers a powerful tool for decision-makers. Health departments, nonprofits, and advocacy groups could use a model like this to allocate mental health resources and track whether interventions are reaching the populations most in need.

There are still many important limitations to acknowledge. First, the dataset relies on self-reported survey data. Individuals may underreport or overreport their mental health status for a variety of reasons including stigma, recall bias, or lack of access to diagnosis. This means the ground truth labels may already be noisy. Second, while one-hot encoding is effective and interpretable, it does not allow the model to learn relationships between categories (e.g., income levels or racial identities with similar contexts). Lastly, the model assumes that past patterns of depression prevalence will hold in the future, which may not be true in the face of large-scale disruptions.

Looking forward, there are several avenues for improvement. One immediate next step is to experiment with learned embeddings for categorical variables, which could improve performance by allowing the model to understand proximity between categories (e.g., adjacent income brackets or neighboring states). Another step that could be taken is to integrate external features, such as unemployment rates, healthcare access data, or education funding, which may explain additional variance in depression prevalence. Finally, deploying the model as part of an interactive dashboard could make the insights more accessible to public health professionals and enable real-time updates as new BRFSS data becomes available. This would be crucial if the model was to be used to find these demographic groups in the future.

In conclusion, this project lays the groundwork for a robust, data-driven approach to identifying population-level mental health risk using demographic data. With further refinement and thoughtful application, it has the potential to play a role in shaping a more targeted, equitable, and effective mental health policy.

Appendix A: Visualizations

Figure A1 – Model Architecture

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
ImprovedDepressionPredictor              [64, 1]                   --
├─Sequential: 1-1                        [64, 1]                   --
│    └─Linear: 2-1                       [64, 512]                 51,200
│    └─ReLU: 2-2                         [64, 512]                 --
│    └─BatchNorm1d: 2-3                  [64, 512]                 1,024
│    └─Dropout: 2-4                      [64, 512]                 --
│    └─Linear: 2-5                       [64, 256]                 131,328
│    └─ReLU: 2-6                         [64, 256]                 --
│    └─BatchNorm1d: 2-7                  [64, 256]                 512
│    └─Dropout: 2-8                      [64, 256]                 --
│    └─Linear: 2-9                       [64, 64]                  16,448
│    └─ReLU: 2-10                        [64, 64]                  --
│    └─Linear: 2-11                      [64, 1]                   65
==========================================================================================
Total params: 200,577
Trainable params: 200,577
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 12.84
```

Figure A2 – Comparison of Model Performance Metrics

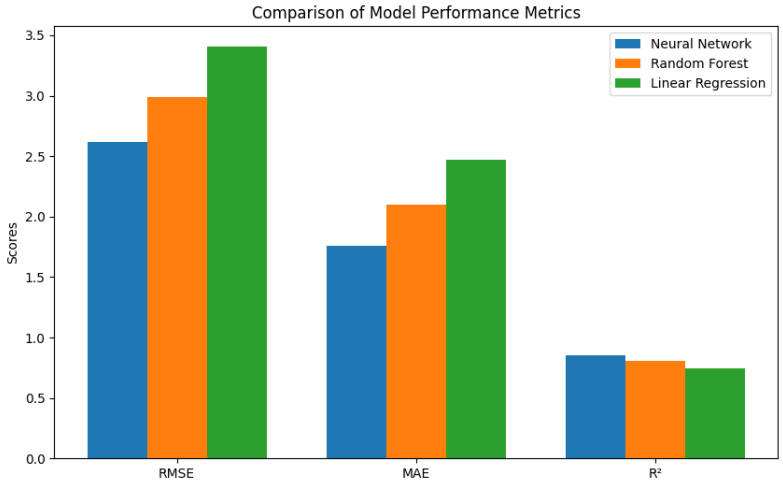


Figure A3 – Predicted vs. Actual Scatterplot

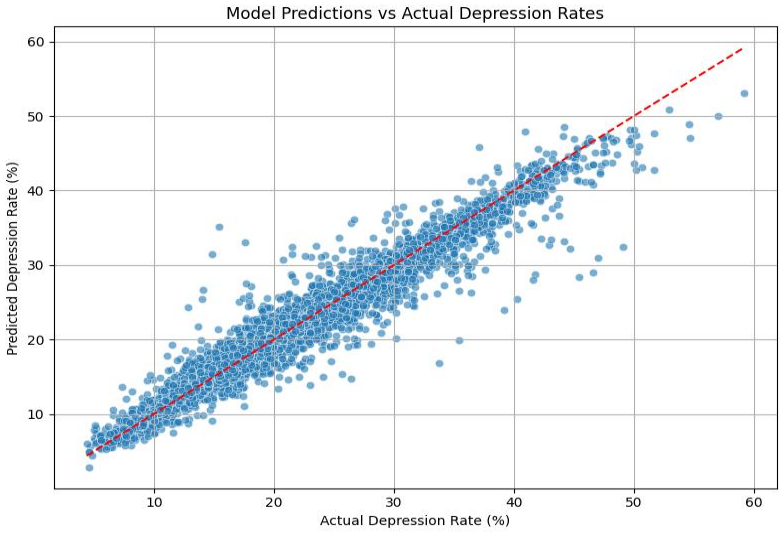


Figure A4 – Top 10 At-Risk Demographic Subgroups

	Year	Locationabbr	Demographic	Predicted_Depression
13879	2022	NH	Household Income - Less than $15,000	53.137508
7783	2017	NH	Household Income - Less than $15,000	50.928761
3054	2013	NH	Race/Ethnicity - Multiracial, non-Hispanic	50.070442
13658	2022	ME	Household Income - Less than $15,000	49.278641
13001	2021	VT	Household Income - Less than $15,000	48.892654
15656	2023	WV	Household Income - Less than $15,000	48.598751
7578	2017	ME	Household Income - Less than $15,000	48.197201
14948	2023	ME	Household Income - Less than $15,000	48.168423
13882	2022	NH	Race/Ethnicity - Multiracial, non-Hispanic	47.940617
5410	2015	NH	Household Income - Less than $15,000	47.729649

Figure A5 – Yearly Depression Rate Trends

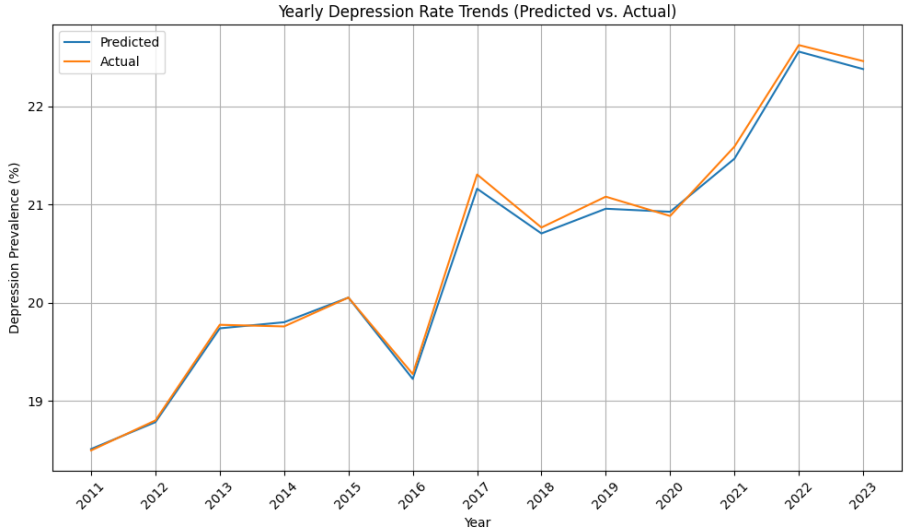
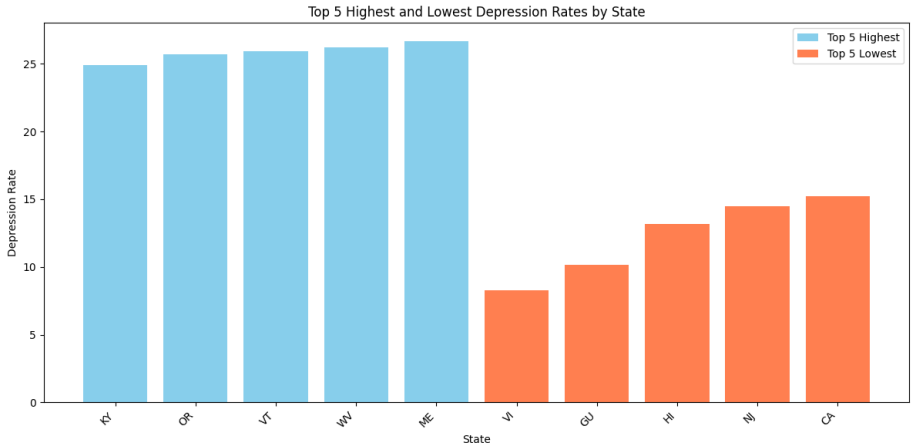


Figure A6 – Highest and Lowest depression rates by state



Appendix B: Code Excerpt – Final Model Definition and Sources

Refer to the Jupyter notebook file for the complete model architecture and grid search implementation. The final model structure is below:

```
class EnhancedDepressionNN(nn.Module):
    def __init__(self, input_dim):
        super(EnhancedDepressionNN, self).__init__()
        self.model = nn.Sequential(
            nn.Linear(input_dim, 512),
            nn.ReLU(),
            nn.BatchNorm1d(512),
            nn.Dropout(0.3),
            nn.Linear(512, 256),
            nn.ReLU(),
            nn.BatchNorm1d(256),
            nn.Dropout(0.2),
            nn.Linear(256, 64),
            nn.ReLU(),
            nn.Linear(64, 1)
        )
    def forward(self, x):
        return self.model(x)
```

Sources:

Centers for Disease Control and Prevention (CDC).
Behavioral Risk Factor Surveillance System (BRFSS).
U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
Available at: https://www.cdc.gov/brfss

Stack Overflow Community.
Discussions and examples related to implementing machine learning models with scikit-learn.
Retrieved from: https://stackoverflow.com/questions/tagged/scikit-learn

Stack Overflow Community.
Code examples and troubleshooting related to building and training neural networks with PyTorch.
Retrieved from: https://stackoverflow.com/questions/tagged/pytorch

Data source and IPYNB File:

Behavioral_Risk_Factor_Surveillance_System__BRFSS__Prevalence_Data__2011_to_present__20250421 (1).csv

Data5610FinalProject (1).ipynb

Note that AI was used in code generation and concept organization