Using ChatGPT for Model Optimization and Data Analysis

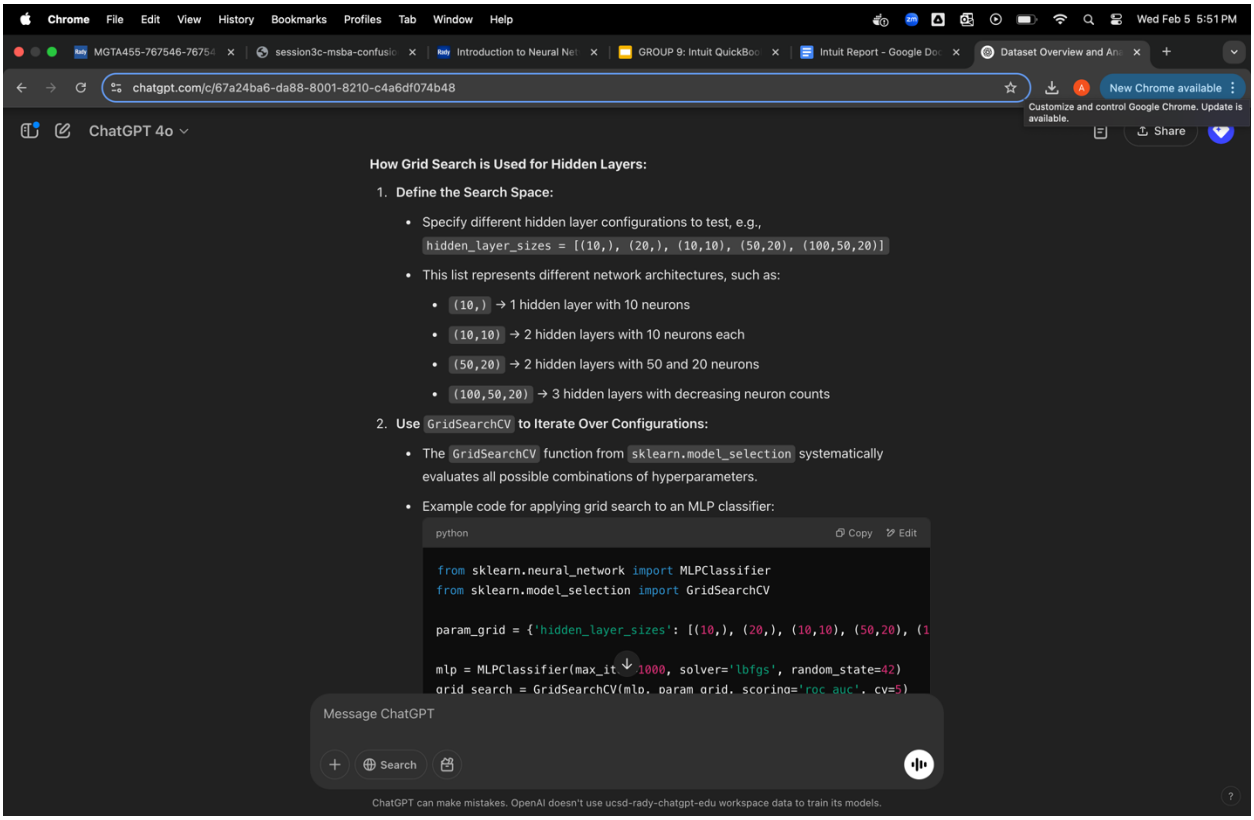
Specific Examples of Prompts Used:
- 'How is Grid Search used for tuning hidden layers in a neural network?'
- 'Provide Python code to implement Grid Search for MLP hidden layers in sklearn.'
'How do I scale and unscaled selected columns in a dataset?'
' Can you extend the code to provide a comprehensive list of pairs up to (5,5)?'

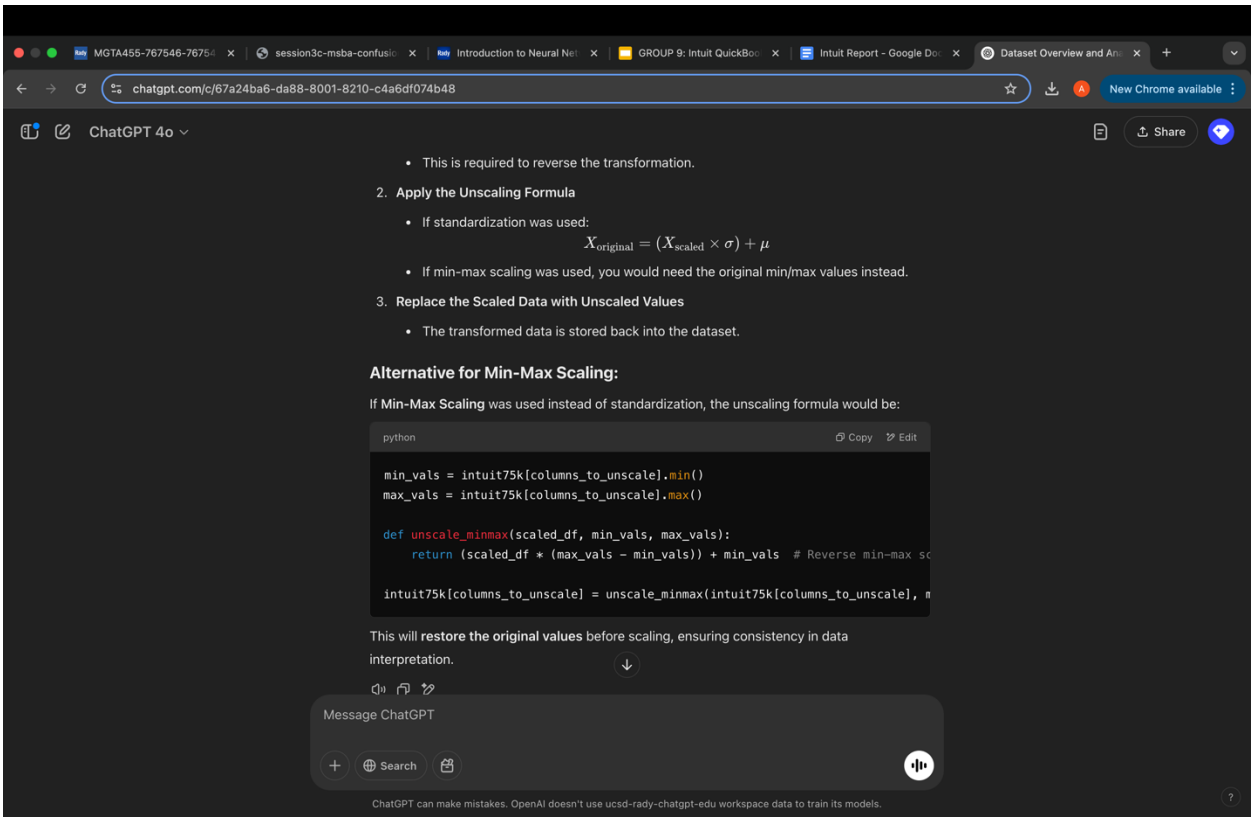
How AI Responses Helped or Hindered Understanding
- **Helped:** The AI provided clear, well-structured explanations and refined our responses, which improved readability and efficiency.
- **Hindered:** At times, the AI offered overly simplified responses or did not fully understand our specific context without additional clarification.

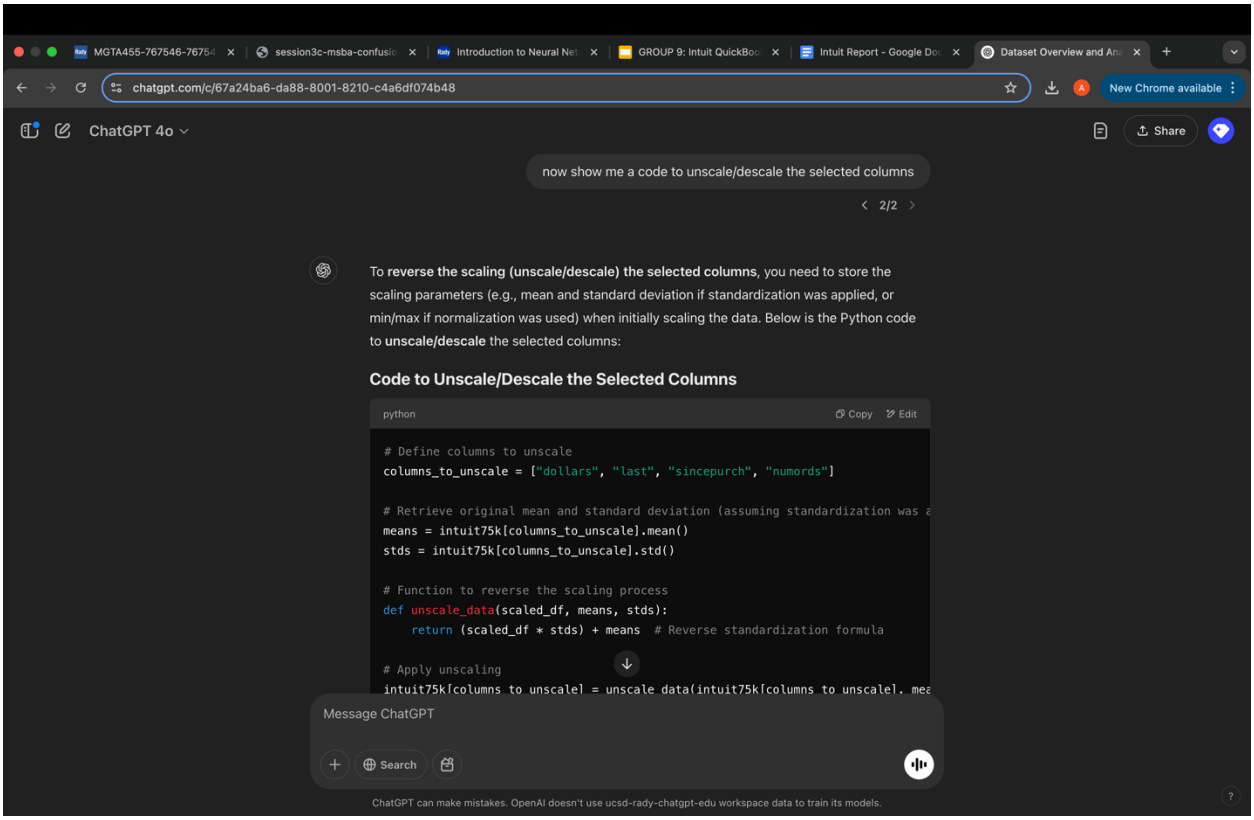
It provides insights into how AUC (Area Under the Curve) is a crucial metric when selecting a model for targeting customers in Wave 2 because it measures how well the model distinguishes between positive and negative cases. Side by side ChatGPT was insightful was its explanations regarding the use of cross validation and grid search for model building and selection, It was also helpful in provided approaches and steps to be taken to analyze projected profit and ROME calculations.

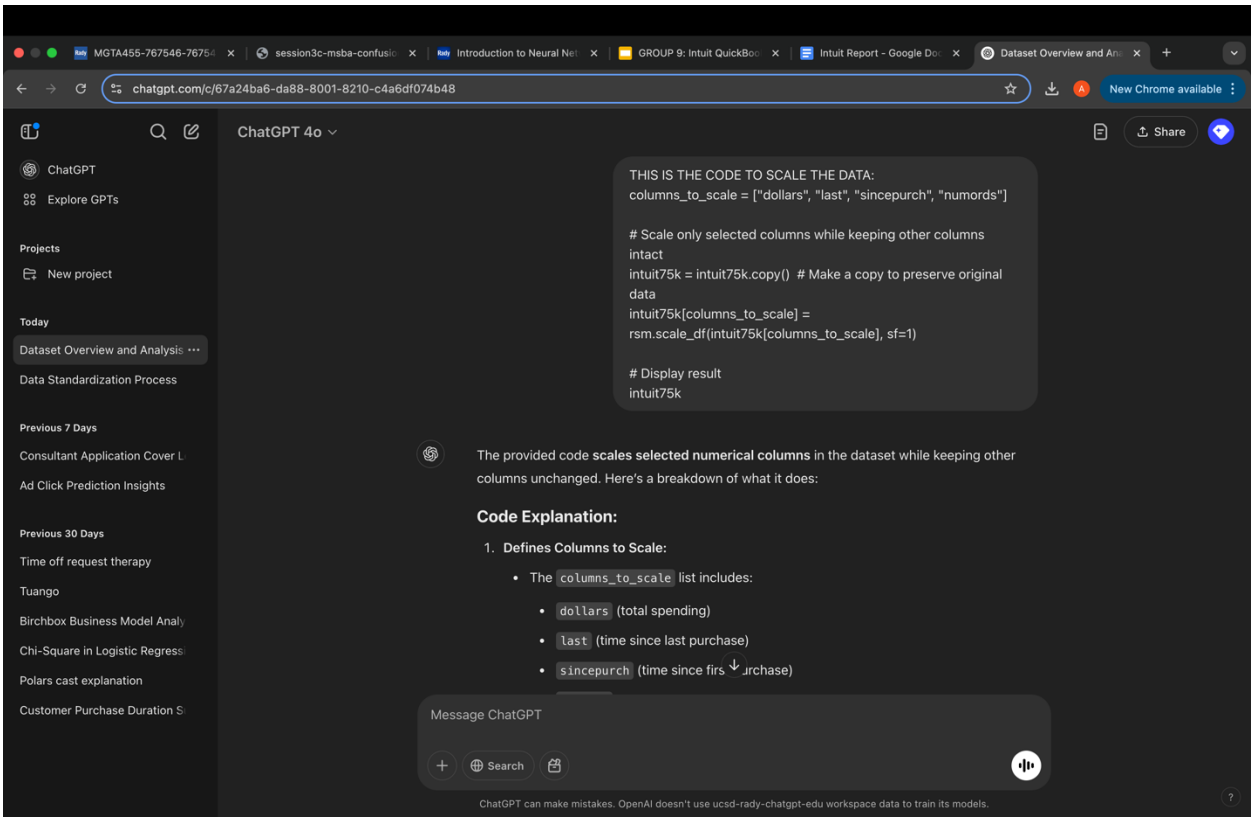
Limitations
Tendency for Overgeneralization: Some responses pulled in irrelevant third-party data, requiring careful verification. ChatGPT kept giving us the wrong prompt and explanations
Also, ChatGPT occasionally provided generic answers that needed refinement. Losing ongoing conversations forced re-explanations, reducing workflow efficiency.
It sometimes suggested solutions that were not optimal or required tweaking to fit our specific needs.

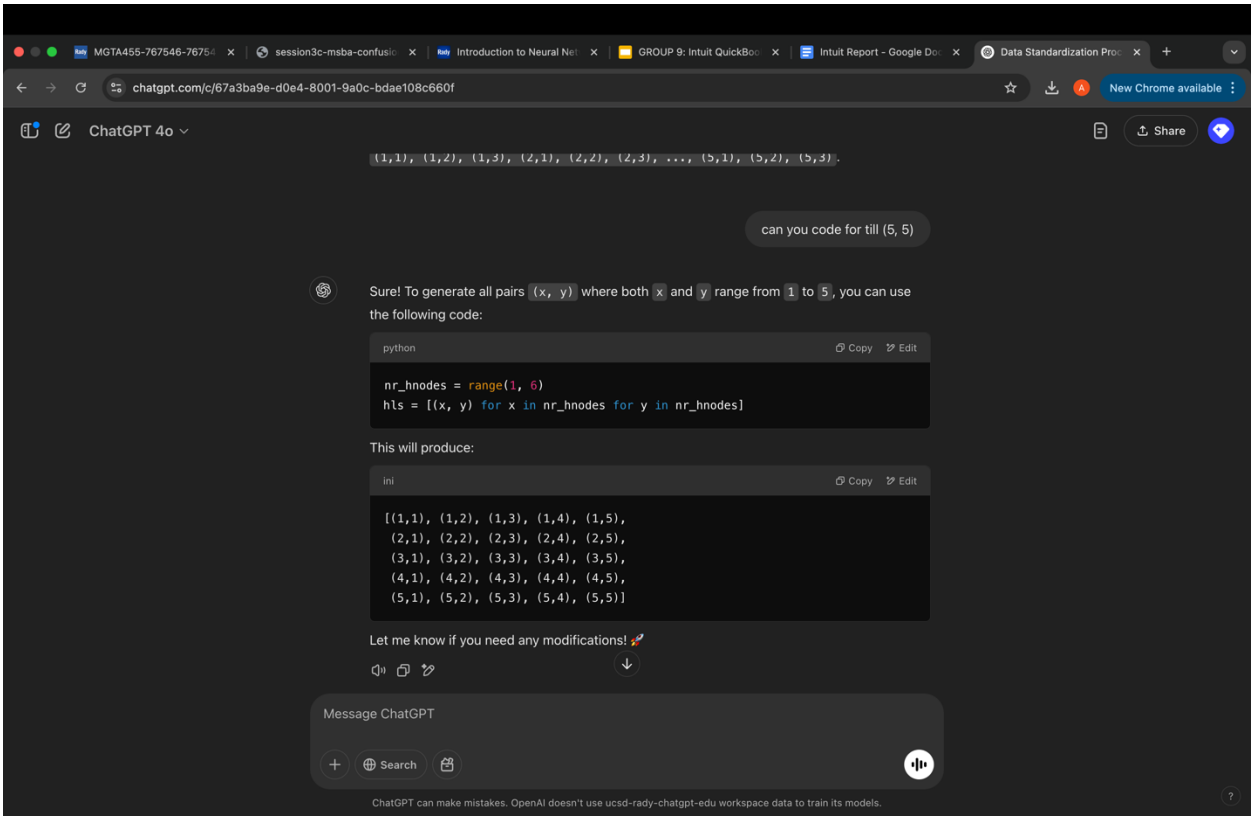
How AI Complemented
Using ChatGPT was a valuable tool in supporting model optimization, data preprocessing, etc. It provided logical explanations and code snippets, enhancing efficiency. However, human oversight was essential to validate responses, refine methodologies, and ensure alignment with business objectives. By using ChatGPT as an **assistive tool rather than a complete solution**, it is more useful to broaden understanding while critically evaluating outputs. Moving forward, we plan to refine our approach to prompt engineering and validation to further enhance our learning experience.


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed Feb 5 5:51 PM
chatgpt.com/c/67a24ba6-da88-8001-8210-c4a6df074b48
New Chrome available
Customize and control Google Chrome. Update is available.
ChatGPT 4o
Share
How Grid Search is Used for Hidden Layers:
1. Define the Search Space:
• Specify different hidden layer configurations to test, e.g., hidden_layer_sizes = [(10,), (20,), (10,10), (50,20), (100,50,20)]
• This list represents different network architectures, such as:
• (10,) → 1 hidden layer with 10 neurons
• (10,10) → 2 hidden layers with 10 neurons each
• (50,20) → 2 hidden layers with 50 and 20 neurons
• (100,50,20) → 3 hidden layers with decreasing neuron counts
2. Use GridSearchCV to Iterate Over Configurations:
• The GridSearchCV function from sklearn.model_selection systematically evaluates all possible combinations of hyperparameters.
• Example code for applying grid search to an MLP classifier:
python
Copy
Edit
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import GridSearchCV
param_grid = {'hidden_layer_sizes': [(10,), (20,), (10,10), (50,20), (1
mlp = MLPClassifier(max_it 1000, solver='lbfgs', random_state=42)
grid_search = GridSearchCV(mlp, param_grid, scoring='roc_auc', cv=5)
Message ChatGPT
Search
ChatGPT can make mistakes. OpenAI doesn't use ucsd-rady-chatgpt-edu workspace data to train its models.


chatgpt.com/c/67a24ba6-da88-8001-8210-c4a6df074b48
New Chrome available
ChatGPT 4o
Share
• This is required to reverse the transformation.
2. Apply the Unscaling Formula
• If standardization was used:
$$X_{\text{original}} = (X_{\text{scaled}} \times \sigma) + \mu$$
• If min-max scaling was used, you would need the original min/max values instead.
3. Replace the Scaled Data with Unscaled Values
• The transformed data is stored back into the dataset.
Alternative for Min-Max Scaling:
If Min-Max Scaling was used instead of standardization, the unscaling formula would be:
python
Copy
Edit
min_vals = intuit75k[columns_to_unscale].min()
max_vals = intuit75k[columns_to_unscale].max()
def unscale_minmax(scaled_df, min_vals, max_vals):
return (scaled_df * (max_vals - min_vals)) + min_vals # Reverse min-max sc
intuit75k[columns_to_unscale] = unscale_minmax(intuit75k[columns_to_unscale], m
This will restore the original values before scaling, ensuring consistency in data interpretation.
Message ChatGPT
Search
ChatGPT can make mistakes. OpenAI doesn't use ucsd-rady-chatgpt-edu workspace data to train its models.


chatgpt.com/c/67a24ba6-da88-8001-8210-c4a6df074b48
New Chrome available
ChatGPT 4o
Share
now show me a code to unscale/descale the selected columns
2/2
To reverse the scaling (unscale/descale) the selected columns, you need to store the scaling parameters (e.g., mean and standard deviation if standardization was applied, or min/max if normalization was used) when initially scaling the data. Below is the Python code to unscale/descale the selected columns:
Code to Unscale/Descale the Selected Columns
python
Copy
Edit
# Define columns to unscale
columns_to_unscale = ["dollars", "last", "sincepurch", "numords"]
# Retrieve original mean and standard deviation (assuming standardization was
means = intuit75k[columns_to_unscale].mean()
stds = intuit75k[columns_to_unscale].std()
# Function to reverse the scaling process
def unscale_data(scaled_df, means, stds):
return (scaled_df * stds) + means # Reverse standardization formula
Apply unscaling
Message ChatGPT
Search
ChatGPT can make mistakes. OpenAI doesn't use ucsd-rady-chatgpt-edu workspace data to train its models.


chatgpt.com/c/67a24ba6-da88-8001-8210-c4a6df074b48
New Chrome available
ChatGPT 4o
Share
ChatGPT
Explore GPTs
Projects
New project
Today
Dataset Overview and Analysis
Data Standardization Process
Previous 7 Days
Consultant Application Cover L
Ad Click Prediction Insights
Previous 30 Days
Time off request therapy
Tuango
Birchbox Business Model Analy
Chi-Square in Logistic Regress
Polars cast explanation
Customer Purchase Duration S
THIS IS THE CODE TO SCALE THE DATA:
columns_to_scale = ["dollars", "last", "sincepurch", "numords"]
# Scale only selected columns while keeping other columns intact
intuit75k = intuit75k.copy() # Make a copy to preserve original data
intuit75k[columns_to_scale] = rsm.scale_df(intuit75k[columns_to_scale], sf=1)
# Display result
intuit75k
The provided code scales selected numerical columns in the dataset while keeping other columns unchanged. Here's a breakdown of what it does:
Code Explanation:
1. Defines Columns to Scale:
The columns_to_scale list includes:
dollars (total spending)
last (time since last purchase)
sincepurch (time since firs rchase)
Message ChatGPT
Search
ChatGPT can make mistakes. OpenAI doesn't use ucsd-rady-chatgpt-edu workspace data to train its models.


MGTA455-767546-76754
session3c-msba-confusio
Introduction to Neural Net
GROUP 9: Intuit QuickBoo
Intuit Report - Google Doc
Data Standardization Proc
chatgpt.com/c/67a3ba9e-d0e4-8001-9a0c-bdae108c660f
New Chrome available
ChatGPT 4o
Share
(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), ..., (5,1), (5,2), (5,3).
can you code for till (5, 5)
Sure! To generate all pairs (x, y) where both x and y range from 1 to 5, you can use the following code:
python
Copy
Edit
nr_hnodes = range(1, 6)
hls = [(x, y) for x in nr_hnodes for y in nr_hnodes]
This will produce:
ini
Copy
Edit
[(1,1), (1,2), (1,3), (1,4), (1,5),
(2,1), (2,2), (2,3), (2,4), (2,5),
(3,1), (3,2), (3,3), (3,4), (3,5),
(4,1), (4,2), (4,3), (4,4), (4,5),
(5,1), (5,2), (5,3), (5,4), (5,5)]
Let me know if you need any modifications!
Message ChatGPT
Search
ChatGPT can make mistakes. OpenAI doesn't use ucsd-rady-chatgpt-edu workspace data to train its models.