NAME: Mannuru. Nisha Prabhu

REG No: 2327234

COURSE: Machine Learning Algorithms - I

TOPIC: Domain-specific model building

Domain - Specific Model Building-Technology

CIA-I

Problem Statement:

To optimize client satisfaction and operational efficiency in interior design projects through data-driven insights.

Objectives:

1. Predict client satisfaction using demographic and project-related data.
2. Optimize project costs while maintaining quality standards.
3. Enhance client experience and service delivery based on feedback and analytics.

DATA DICTIONARY:

Column Name	Data Type	Description	Example Value
project_id	Integer	Unique identifier for the project	1
client_age	Float	Age of the client	63
client_gender	String	Gender of the client	Male
project_budget	Float	Budget allocated for the project	64042
project_duration	Integer	Duration of the project (in weeks)	16
num_rooms	Integer	Number of rooms in the project	7
design_style	String	Design style chosen for the project	Eclectic
location	String	Location of the project	Urban
client_satisfaction	Float	Client satisfaction rating (1-10)	9
designer_experience	Float	Experience of the designer (in years)	3
renovation_cost	Integer	Cost of renovation (if any)	3364
project_completion_time	Integer	Time taken to complete the project	22
num_meetings	Integer	Number of meetings with the client	1
client_feedback	String	Feedback from the client	Negative
material_availability	String	Availability of materials	Not available
designer_rating	Integer	Rating of the designer (1-5)	3
project_rating	Integer	Overall project rating (1-5)	5
design_approval_time	Integer	Time taken for design approval	2
issues_encountered	String	Issues encountered during the project	No
total_cost	Float	Total cost of the project	197451

Variables:

- **client_satisfaction:** The target variable, representing the level of satisfaction of the client (on a scale, possibly 1-10 or another range).
- **client_age:** The age of the client.
- **designer_experience:** The years of experience of the designer handling the project.
- **num_designers:** The number of designers involved in the project.
- **furniture_cost:** The cost incurred on furniture for the project.
- **lighting_cost:** The cost incurred on lighting for the project.
- **decor_cost:** The cost incurred on decor for the project.
- **renovation_cost:** The cost incurred on renovation for the project.
- **project_completion_time:** The time taken to complete the project.
- **num_meetings:** The number of meetings held with the client.

OUTPUTS

Multiple Linear Regression:

DATA AFTER SPLITTING 80% TRAINING 20 % TESTING

```
Call:
lm(formula = client_satisfaction ~ client_age + designer_experience +
    num_designers + furniture_cost + lighting_cost + decor_cost +
    renovation_cost + project_completion_time + num_meetings,
    data = train_data)

Residuals:
   Min     1Q Median     3Q    Max
-5.056 -2.529  0.227  2.484  4.832

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.60433    0.09935  56.410   <2e-16 ***
client_age               0.06867    0.09986   0.688    0.492
designer_experience      0.06640    0.09971   0.666    0.506
num_designers           -0.09577    0.09975  -0.960    0.337
furniture_cost          -0.05729    0.10001  -0.573    0.567
lighting_cost            0.03488    0.10054   0.347    0.729
decor_cost               0.03942    0.09891   0.399    0.690
renovation_cost          0.09943    0.09961   0.998    0.318
project_completion_time -0.01414    0.10032  -0.141    0.888
num_meetings             0.08525    0.10036   0.849    0.396
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.887 on 836 degrees of freedom
Multiple R-squared:  0.004746,  Adjusted R-squared:  -0.005968
F-statistic: 0.443 on 9 and 836 DF,  p-value: 0.9118
```

```
# Predictions on the test set

> predictions <- predict(model, test_data)

> # Model evaluation

> actuals <- test_data$client_satisfaction

> mse <- mean((predictions - actuals)^2)

> rmse <- sqrt(mse)

> mae <- mean(abs(predictions - actuals))

> # Print the evaluation metrics

> cat("Mean Squared Error (MSE): ", mse, "\n")

Mean Squared Error (MSE):  7.916308

> cat("Root Mean Squared Error (RMSE): ", rmse, "\n")

Root Mean Squared Error (RMSE):  2.813593

> cat("Mean Absolute Error (MAE): ", mae, "\n")

Mean Absolute Error (MAE):  2.426947
```

RIDGE REGRESSION

```
> # Print the evaluation metrics
> cat("Mean Squared Error (MSE): ", mse, "\n")
Mean Squared Error (MSE):  7.745321
> cat("Root Mean Squared Error (RMSE): ", rmse, "\n")
Root Mean Squared Error (RMSE):  2.783042
> cat("Mean Absolute Error (MAE): ", mae, "\n")
Mean Absolute Error (MAE):  2.40566
```

```
# Fit the Ridge Regression model

> ridge_model <- glmnet(x_train, y_train, alpha = 0)

> # Cross-validation to find the optimal lambda

> cv_ridge <- cv.glmnet(x_train, y_train, alpha = 0)

> # Best lambda value
```

```
> best_lambda <- cv_ridge$lambda.min
> cat("Best lambda: ", best_lambda, "\n")
Best lambda:  99.8603
> # Predict on the test set using the best lambda
> ridge_predictions <- predict(ridge_model, s = best_lambda, newx = x_test)
> # Model evaluation
> mse <- mean((ridge_predictions - y_test)^2)
> rmse <- sqrt(mse)
> mae <- mean(abs(ridge_predictions - y_test))
> # Print the evaluation metrics
> cat("Mean Squared Error (MSE): ", mse, "\n")
Mean Squared Error (MSE):  7.745321
> cat("Root Mean Squared Error (RMSE): ", rmse, "\n")
Root Mean Squared Error (RMSE):  2.783042
> cat("Mean Absolute Error (MAE): ", mae, "\n")
Mean Absolute Error (MAE):  2.40566
```

LASSO REGRESSION

```
> # Print the evaluation metrics
> cat("Mean Squared Error (MSE): ", mse, "\n")
Mean Squared Error (MSE):  7.745321
> cat("Root Mean Squared Error (RMSE): ", rmse, "\n")
Root Mean Squared Error (RMSE):  2.783042
> cat("Mean Absolute Error (MAE): ", mae, "\n")
Mean Absolute Error (MAE):  2.40566
```

```
# Best lambda value
> best_lambda <- cv_lasso$lambda.min
> cat("Best lambda: ", best_lambda, "\n")
Best lambda:  0.0998603
> # Predict on the test set using the best lambda
> lasso_predictions <- predict(lasso_model, s = best_lambda, newx = x_test)
```

```
> # Model evaluation
> mse <- mean((lasso_predictions - y_test)^2)
> rmse <- sqrt(mse)
> mae <- mean(abs(lasso_predictions - y_test))
> # Print the evaluation metrics
> cat("Mean Squared Error (MSE): ", mse, "\n")
Mean Squared Error (MSE):  7.745321
> cat("Root Mean Squared Error (RMSE): ", rmse, "\n")
Root Mean Squared Error (RMSE):  2.783042
> cat("Mean Absolute Error (MAE): ", mae, "\n")
Mean Absolute Error (MAE):  2.40566
```

2. Model Output

Explanation of the Parameters

Intercept ($\beta 0$): The expected value of client satisfaction when all predictors are zero.

Coefficients ($\beta 1, \beta 2, ..., \beta 9$): The change in client satisfaction for a one-unit change in the predictor variable, holding all other variables constant.

Explanation of the Coefficients

From the multiple linear regression model output: None of the coefficients are statistically significant (all p-values are > 0.05), indicating that none of the predictors have a statistically significant effect on client satisfaction in this model.

Model Fit Indices

Model fit indices help assess how well the model describes the observed data. The following are the key indices from the multiple linear regression model and their interpretations:

a. **Residual Standard Error (RSE): 2.887**
   This measures the average amount that the observed values deviate from the model's predicted values. A lower RSE indicates a better fit. In this case, the RSE of 2.887 suggests that, on average, the predicted client satisfaction scores deviate from the actual scores by about 2.887 units.

b. **Multiple R-squared: 0.004746**

This represents the proportion of variance in the dependent variable (client satisfaction) that is explained by the independent variables in the model. An R-squared value of 0.004746 means that approximately 0.47% of the variability in client satisfaction is explained by the model. This is very low, indicating that the model does not explain much of the variation in client satisfaction.

c. **Adjusted R-squared: -0.005968**
The adjusted R-squared adjusts the R-squared value for the number of predictors in the model. It penalizes the addition of non-significant predictors. In this case, the adjusted R-squared is negative, which can happen when the model does not perform well and the predictors do not explain the variability in the dependent variable. This suggests that the model might be overfitting the data or that the predictors are not relevant.

d. **F-statistic: 0.443 on 9 and 836 DF**
The F-statistic tests the overall significance of the model. It compares the fit of the model with and without the predictors. A higher F-statistic indicates that the model is a better fit. In this case, the F-statistic of 0.443 with a p-value of 0.9118 indicates that the model is not statistically significant and does not provide a better fit than a model with no predictors.

3. Model Interpretation from the Business Point of View

From a business perspective, the results suggest that the current predictors (client age, designer experience, number of designers, various costs, project completion time, and number of meetings) do not significantly influence client satisfaction. This implies that other factors, not included in this model, might be driving client satisfaction. These could include qualitative aspects such as the client's personal preferences, communication quality, design creativity, or unexpected project delays and issues.

4. Model Evaluation and Diagnostics

Evaluation Metrics for Multiple Linear Regression

Mean Squared Error (MSE): 7.916308

Root Mean Squared Error (RMSE): 2.813593

Mean Absolute Error (MAE): 2.426947

These metrics indicate that the model's predictions deviate significantly from the actual client satisfaction scores. The high RMSE and MAE values suggest that the model does not predict client satisfaction accurately.

Evaluation Metrics for Ridge and Lasso Regression

Both Ridge and Lasso regression models had similar evaluation metrics:

Mean Squared Error (MSE): 7.745321

Root Mean Squared Error (RMSE): 2.783042

Mean Absolute Error (MAE): 2.40566

These metrics indicate that both Ridge and Lasso regression models perform slightly better than the multiple linear regression model. However, the improvements are marginal, suggesting that the chosen predictors may not be highly influential for client satisfaction.

5. A Short Note on Democratizing the Solution

- Democratizing the solution involves making the predictive models and insights accessible to a broader audience within the organization. This can be achieved by:
- Developing intuitive dashboards: Use tools like PowerBI to create interactive dashboards that display real-time insights and predictive analytics.
- Providing training: Ensure that staff members are trained to interpret and utilize the insights from the models.
- Encouraging data-driven decision-making: Promote a culture where decisions are based on data and predictive insights rather than intuition alone.
- Simplifying access to data: Ensure that relevant data is easily accessible to those who need it, with appropriate data governance in place.

6. Conclusion

The analysis aimed to identify key drivers of client satisfaction in interior design projects using multiple linear regression, Ridge regression, and Lasso regression models. The results indicated that the current set of predictors does not significantly influence client satisfaction, as evidenced by low R-squared values and non-significant coefficients. This suggests the need to explore additional variables and possibly non-linear relationships to improve the model's predictive performance.

Future steps could include qualitative research to identify other factors influencing client satisfaction, collecting more comprehensive data, and experimenting with different modeling techniques. Democratizing the solution by making insights accessible and promoting data-driven decision-making within the organization can help leverage the predictive models to enhance client satisfaction and improve business outcomes.This comprehensive approach will help the organization better understand and address the factors affecting client satisfaction, ultimately leading to improved service quality and client relationships.