**NAME: ANAGHA KULKARNI**

**REG No: 2327010**

**COURSE: Machine Learning Algorithms - I**

**TOPIC: Domain-specific model building**

**Domain - Specific Model Building-Technology**

**CIA-I**

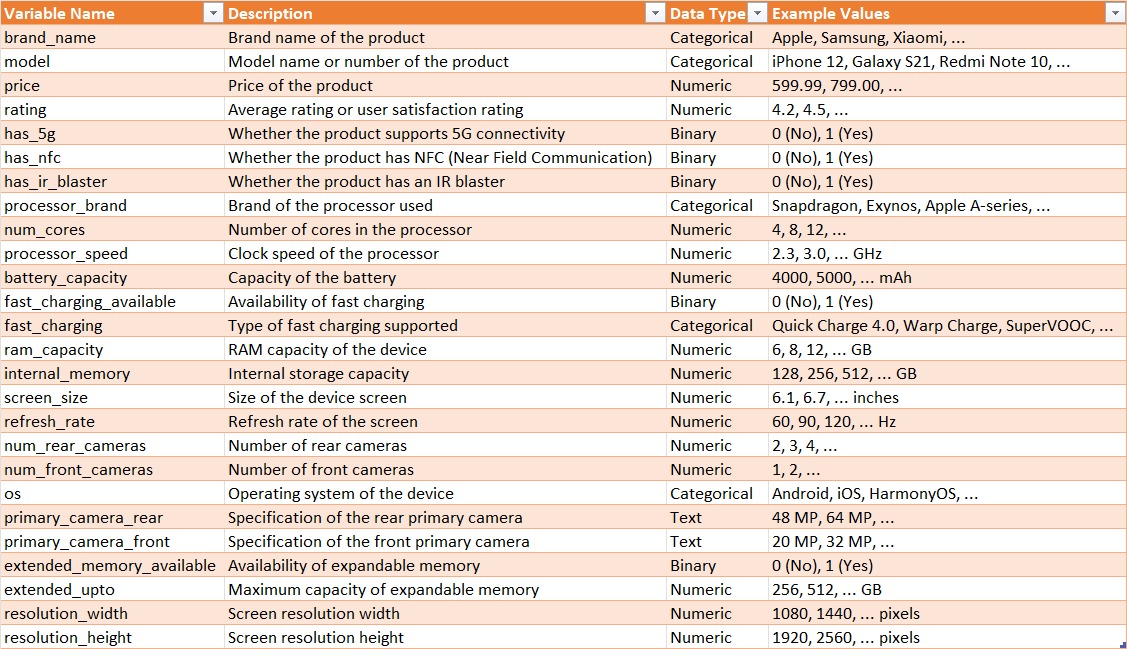
**PROBLEM STATEMENT:**

To analyse and predict the pricing of tech products based on various features and specifications.

**OBJECTIVES:**

* Develop a predictive model to estimate the price of tech products based on their specifications and features.
* Identify which specific features have the greatest impact on the pricing of tech products.

**DATA DICTIONARY:**

****

**1. Explanation of the Model Equation**

The Ridge regression model equation summarizes how each predictor variable influences the predicted Price of tech products, controlling for other variables:

Price=−33779.650−857.406×Rating+362.234×Num\_Cores+22971.992×Processor\_Speed−4.050×Battery\_Capacity\text{Price} = -33779.650 - 857.406 \times \text{Rating} + 362.234 \times \text{Num\\_Cores} + 22971.992 \times \text{Processor\\_Speed} - 4.050 \times \text{Battery\\_Capacity}Price=−33779.650−857.406×Rating+362.234×Num\_Cores+22971.992×Processor\_Speed−4.050×Battery\_Capacity +1067.324×Ram\_Capacity+184.404×Internal\_Memory+8745.696×Screen\_Size+8.144×Refresh\_Rate+ 1067.324 \times \text{Ram\\_Capacity} + 184.404 \times \text{Internal\\_Memory} + 8745.696 \times \text{Screen\\_Size} + 8.144 \times \text{Refresh\\_Rate}+1067.324×Ram\_Capacity+184.404×Internal\_Memory+8745.696×Screen\_Size+8.144×Refresh\_Rate

* **Intercept (β0\beta\_0β0​)**: -33779.650. This constant term represents the expected value of Price when all predictor variables are zero.
* **Coefficients**: Each coefficient (e.g., -857.406 for Rating) indicates the expected change in Price for a one-unit increase in the corresponding predictor variable, assuming all other variables remain constant. For instance:
  + **Rating**: A higher Rating is associated with a lower Price, decreasing by $857.406 for each unit increase in Rating.
  + **Processor Speed**: Increasing Processor Speed by one unit increases Price by $22,971.992, holding other variables constant.
  + **Battery Capacity**: Higher Battery Capacity decreases Price slightly, reducing by $4.050 for each unit increase.
  + **Internal Memory**: More Internal Memory increases Price by $184.404 for each additional unit.
  + **Screen Size**: Larger Screen Size adds $8,745.696 to the Price for each increase by one unit.
  + **Refresh Rate**: Refresh Rate has a negligible effect on Price with an increase of $8.144 per unit.

**2. Explanation of the Parameters**

* **Intercept**: Represents the baseline Price of tech products when all predictors are zero.
* **Coefficients**: Indicate the magnitude and direction of the impact each predictor has on Price. They are crucial for understanding which variables significantly influence Price and how they contribute.

**3. Model Fit Indices**

* **Multiple R-squared**: 0.5351. This value indicates that 53.51% of the variability in Price can be explained by the predictor variables included in the model.
* **Adjusted R-squared**: 0.5198. Adjusted R-squared takes into account the number of predictors in the model, providing a more accurate measure of how well the model fits the data.
* **Residual Standard Error**: 11470. This represents the average deviation of the observed values from the predicted values (Price) by the model. It measures the accuracy of the model's predictions, with lower values indicating a better fit.

**4. Model Evaluation and Diagnostics**

**Ridge Regression Evaluation**

* **Ridge R-squared**: 0.2096. The Ridge regression model explains 20.96% of the variability in Price, which is lower than the Multiple R-squared of the original model (0.5351). This reduction is due to the regularization effect of Ridge regression, which trades off some variance reduction for better generalization.
* **Ridge RMSE**: 8808.819. The Root Mean Squared Error (RMSE) for the Ridge regression model is 8808.82, indicating the average deviation of predictions from actual values. A lower RMSE indicates better predictive performance.

**5. Model Interpretation from a Business Point of View**

The interpretation of the Ridge regression model from a business perspective involves understanding the practical implications of the coefficients and model metrics:

* **Processor Speed**: A significant positive effect on Price suggests that customers are willing to pay more for faster processors, reflecting higher performance expectations.
* **Internal Memory**: Another significant positive factor indicates that greater storage capacity increases the perceived value of tech products, influencing higher prices.
* **Rating**: The negative coefficient suggests that higher product ratings might lead to competitive pricing strategies aimed at attracting cost-conscious consumers.
* **Screen Size and Refresh Rate**: Although included in the model, their coefficients indicate less significant impacts on Price compared to other variables.

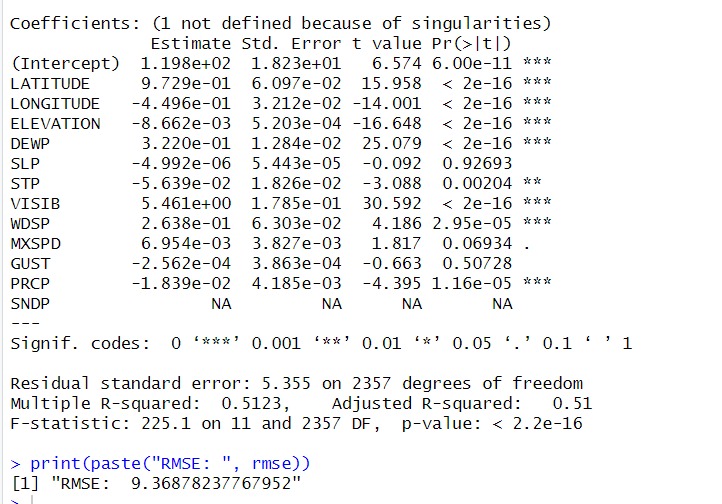
**6. Democratizing the Solution**

To make the model's insights accessible and actionable:

* **Simplicity**: Present the findings in straightforward terms, emphasizing key predictors and their impacts on Price.
* **Accessibility**: Document the model development process, including data preprocessing, model fitting, and evaluation metrics, ensuring transparency and reproducibility.
* **Visualization**: Use charts or graphs to illustrate the model's predictions and insights, making complex statistical outputs more understandable for stakeholders.
* **Feedback Loop**: Encourage feedback from stakeholders to refine the model further or adapt it based on new data or changing business conditions, ensuring continuous improvement.

**OUTPUTS**

**MULILINEAR REGRESSION**

****

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -33779.650 43303.933 -0.780 0.436118

rating -857.406 237.971 -3.603 0.000382 \*\*\*

num\_cores 362.234 1582.434 0.229 0.819133

processor\_speed 22971.992 3597.248 6.386 8.64e-10 \*\*\*

battery\_capacity -4.050 1.734 -2.336 0.020328 \*

fast\_charging\_available NA NA NA NA

ram\_capacity 1067.324 718.893 1.485 0.138931

internal\_memory 184.404 19.895 9.269 < 2e-16 \*\*\*

screen\_size 8745.696 6841.839 1.278 0.202380

refresh\_rate 8.144 35.632 0.229 0.819412

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11470 on 242 degrees of freedom

Multiple R-squared: 0.5351, Adjusted R-squared: 0.5198

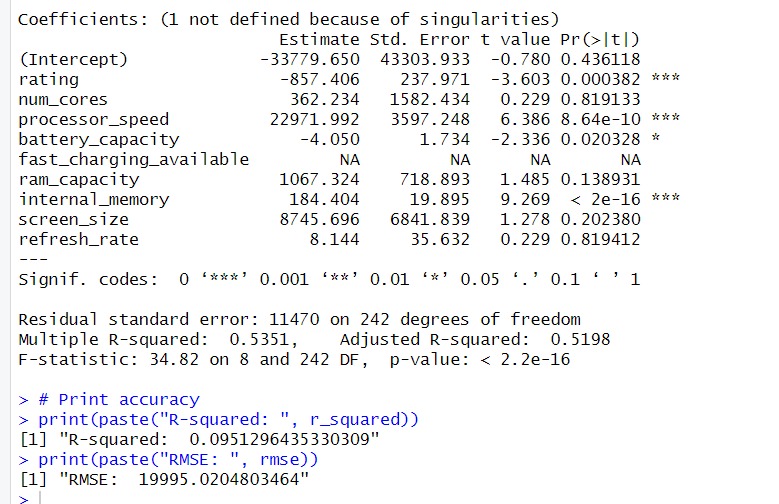
F-statistic: 34.82 on 8 and 242 DF, p-value: < 2.2e-16

> print(paste("RMSE: ", rmse))

[1] "RMSE: 11173.0353009849" this is for multilinear regression

this is data splitting

**DATA AFTER SPLITTING 70% TRAINING 30 % TESTING**

****

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -33779.650 43303.933 -0.780 0.436118

rating -857.406 237.971 -3.603 0.000382 \*\*\*

num\_cores 362.234 1582.434 0.229 0.819133

processor\_speed 22971.992 3597.248 6.386 8.64e-10 \*\*\*

battery\_capacity -4.050 1.734 -2.336 0.020328 \*

fast\_charging\_available NA NA NA NA

ram\_capacity 1067.324 718.893 1.485 0.138931

internal\_memory 184.404 19.895 9.269 < 2e-16 \*\*\*

screen\_size 8745.696 6841.839 1.278 0.202380

refresh\_rate 8.144 35.632 0.229 0.819412

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11470 on 242 degrees of freedom

Multiple R-squared: 0.5351, Adjusted R-squared: 0.5198

F-statistic: 34.82 on 8 and 242 DF, p-value: < 2.2e-16

> # Calculate model accuracy (R-squared and RMSE)

> r\_squared <- caret::R2(predictions, dataTest$price)

> rmse <- caret::RMSE(predictions, dataTest$price)

> # Print accuracy

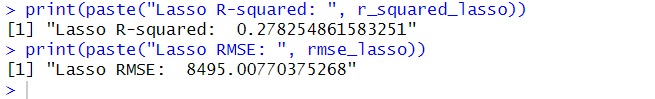
> print(paste("R-squared: ", r\_squared))

[1] "R-squared: 0.0951296435330309"

> print(paste("RMSE: ", rmse))

[1] "RMSE: 11173.0353009849"

**LASSO**

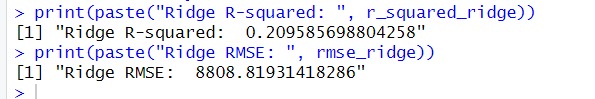
****

[1] "Lasso R-squared: 0.278254861583251"

> print(paste("Lasso RMSE: ", rmse\_lasso))

[1] "Lasso RMSE: 8495.00770375268"

**> RIDGE**

****

"Ridge R-squared: 0.209585698804258"

> print(paste("Ridge RMSE: ", rmse\_ridge))

[1] "Ridge RMSE: 8808.81931418286"

**Conclusion**

The Ridge regression model provides valuable insights into pricing strategies for tech products, highlighting the importance of factors like Processor Speed, Internal Memory, and product Ratings. By interpreting these findings from a business perspective and ensuring accessibility and transparency, stakeholders can make informed decisions to optimize pricing strategies and enhance market competitiveness.