Theran Meadows

Project 2

10 Questions

1. **What were the most significant factors determining survival on the Titanic?**

   Factors such as passenger class (Pclass), sex, and age played significant roles. Historical records and analysis often show that women and children had higher survival rates, and passengers in higher classes were more likely to survive due to better access to lifeboats.

2. **How did you handle missing data in your analysis, especially for variables like Age and Cabin?**

   Missing data were addressed through techniques like imputation. For age, common methods include using the median or mean age, whereas for cabin, due to the high percentage of missing values, it was dropped though having it derived into a simpler form such as 'Cabin Known' vs 'Cabin Unknown' could have been possible.

3. **Can you explain how you chose the model for your analysis? Why did you prefer it over others?**

   The choice of model could be based on its suitability for binary classification problems and the dataset size. Logistic regression is a common choice due to its interpretability and efficiency. More complex models like random forests or support vector machines might be chosen for their higher accuracy in more complex scenarios.

4. **Were there other performance metrics of your predictive model?**

Performance metrics likely include accuracy, precision, recall, and the F1-score. These metrics help evaluate the model's ability to correctly predict survival, balance between sensitivity and specificity, and the harmonic mean of precision and recall, respectively.

5. **How did the socio-economic status (represented by the passenger class) influence the survival rates?**

Socio-economic status, indicated by the passenger class, significantly influenced survival chances. Higher-class passengers often had better access to lifeboats and emergency resources, leading to higher survival rates.

6. **Did the embarkation point affect the survival chances, according to your model?**

The embarkation point might show some influence due to socio-economic factors related to the demographics of passengers boarding at different ports. However, its impact is generally less significant compared to other variables like sex and class.

7. **How did you ensure that your model was not overfitting the training data?**

Techniques such as cross-validation, where the data is split into multiple subsets to validate the model against different parts of the dataset, help prevent overfitting. Regularization methods in model training also reduce the risk of overfitting by penalizing overly complex models.

8. **Could you expand on how feature engineering influenced your model's performance?**

Feature engineering, such as creating new features like family size from SibSp and Parch, or extracting titles from names, could help the model by introducing new relevant information or simplifying existing information, thus potentially increasing predictive accuracy.

9. **Are there ethical considerations in using demographic data (like sex and age) for predictive modeling in this context?**

Ethical considerations arise when using demographic data, as it involves sensitive information. In historical analysis like the Titanic, it's primarily for understanding disparities in survival rather than decision-making. However, in modern applications, such usage requires careful ethical review to avoid reinforcing biases.

It should also be noted that this disaster happened over 100 years ago. The last survivor died in 2009 at the age of 97. She was 3 months old when she boarded with her family.

10. **What further data would you like to have to improve the predictive accuracy of your model?**

Additional data like the physical health of passengers, exact location on the ship at the time of the disaster, or detailed crew actions during the evacuation could provide deeper insights and improve model accuracy by highlighting other survival factors.