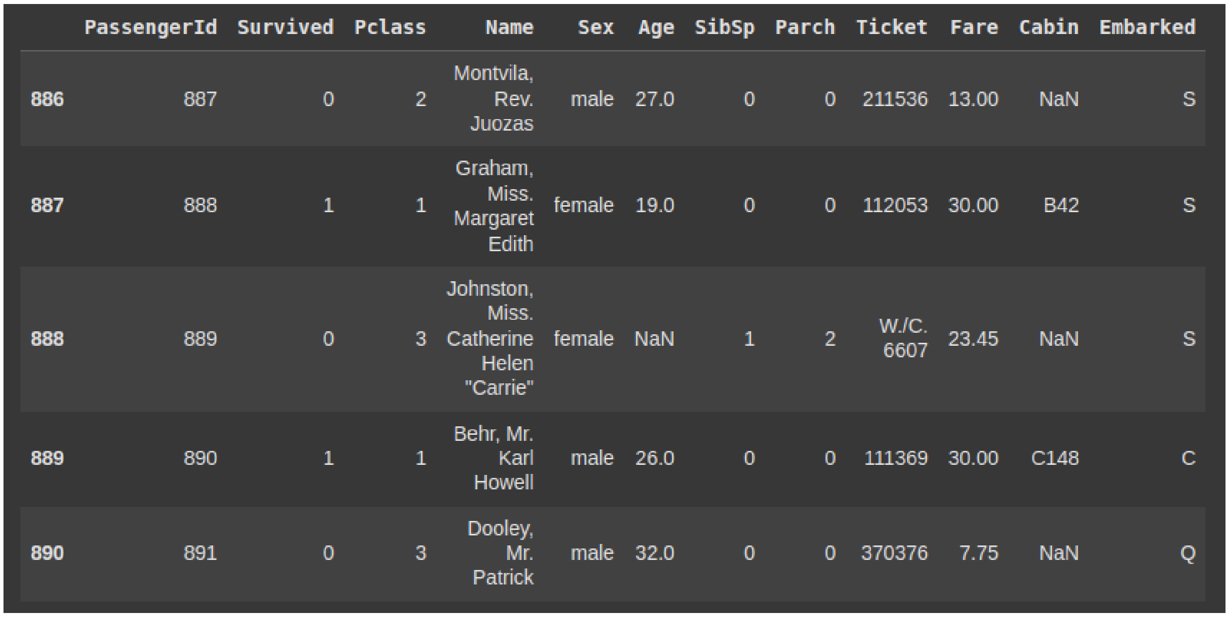
Ethical Analysis of the Dataset and Model

Introduction

One well-known machine learning resource that offers insightful information for forecasting is the Titanic dataset. But it also raises significant ethical questions about biases, justice, and the difficulty of striking a balance between justice and accuracy in model results.



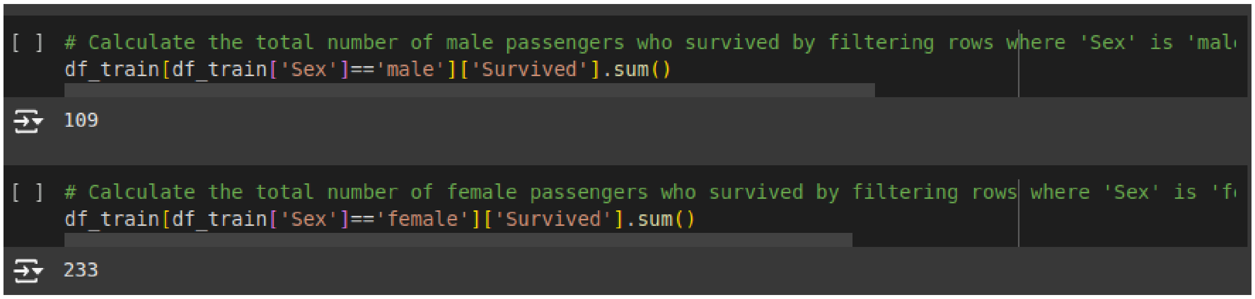
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

Identification of Biases in the Dataset

Historical societal norms and systemic injustices that had a major influence on survival rates are demonstrated by the Titanic dataset. When predicting survival, variables like age, gender, and passenger class (Pclass) are crucial. Women, for example, had greater survival rates, primarily because of societal norms that put their safety first during evacuations. This has less to do with innate survival advantage and more to do with social interactions.

First-class passengers had a significantly higher chance of surviving than those in second or third class, further demonstrating socioeconomic disparities. Higher classes disproportionately profited from access to lifeboats and rescue efforts, underscoring the privileges that come with wealth and social status.

The results of survival were likewise influenced by age. During evacuation operations, younger passengers, especially children were frequently prioritized. The dataset was further impacted by an age-related bias in survival rates brought about by this emphasis on young. These historical differences show that social norms, not practical survival capabilities, influenced survival results. Therefore, distorted predictions may occur from models unintentionally learning and reinforcing these biases when trained on this dataset.



```
[ ] # Calculate the total number of male passengers who survived by filtering rows where 'Sex' is 'mal
    df_train[df_train['Sex']=='male']['Survived'].sum()

    109

[ ] # Calculate the total number of female passengers who survived by filtering rows where 'Sex' is 'f
    df_train[df_train['Sex']=='female']['Survived'].sum()

    233
```

Impact of Bias on Fairness of Model Predictions

The Titanic dataset contains biases that machine learning models may inadvertently perpetuate when they are trained on it. Social dynamics that favored women, younger travelers, and passengers from higher socioeconomic strata are seen in the dataset. As a result, these models could unjustly give these groups preference in modern applications like resource allocation or evacuation planning.

A model trained on this data, for example, would forecast that female travelers, younger people, or first-class passengers will have better survival rates. These forecasts are based on past social preferences rather than real survival considerations. Particularly if the approach is applied in circumstances when fair treatment is essential, this biased prioritizing compromises justice and may result in immoral choices. Such biased forecasts are incompatible with contemporary ideals of equality and fairness, underscoring the risks of depending on historical data without taking into account its inherent flaws.

Trade-offs Between Accuracy and Fairness

Finding the ideal balance between fairness and accuracy is a major difficulty in machine learning. Using factors like Pclass and Sex, which are strongly associated with survival outcomes, is frequently necessary when concentrating on accuracy. Although this tactic can improve performance indicators, it also produces biased projections and reinforces past biases.

However, in order to reduce prejudice, highlighting fairness could necessitate reweighting or removing sensitive elements. Even though this can result in more equitable forecasts, accuracy usually suffers as a result of the model perhaps missing out on important information. This leads to a moral conundrum: should we put performance maximization or maintaining equity first?

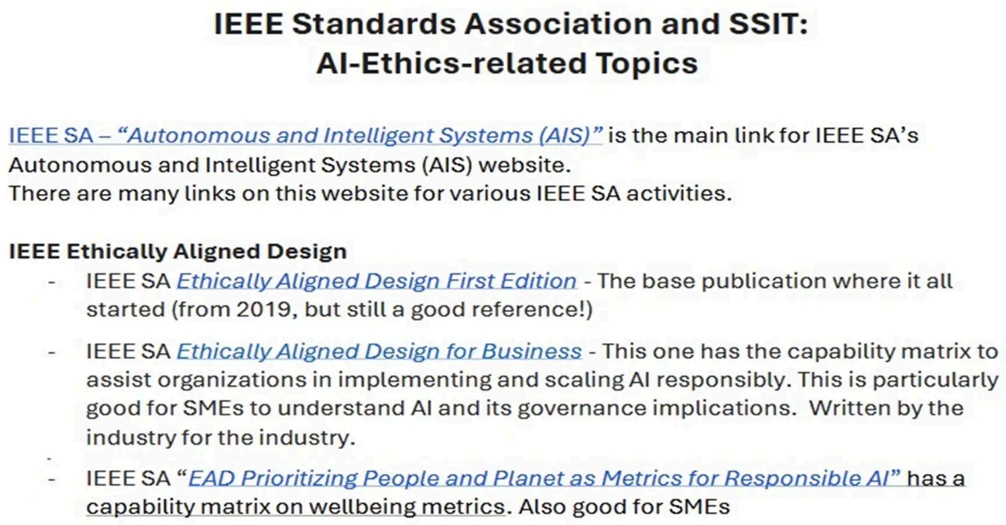
Accuracy alone may demonstrate the potential of predictive models in exploratory or instructive contexts. Fairness should, however, come first in real-world applications to stop historical injustices from happening again. Intentional efforts to lessen biases while preserving respectable prediction performance are necessary to achieve ethical decision-making.

Real-World Implications of Biased Models

There can be serious repercussions when biased models developed on the Titanic dataset are applied in practical settings. These models might unjustly benefit particular groups based on gender, age, or socioeconomic status, for instance, if they were applied to evacuation planning or resource distribution. This might undermine confidence in AI systems and exacerbate already-existing disparities.

For example, a biased model can put younger or perceived wealthier passengers first for evacuation, putting older or lower socioeconomic group members at a disadvantage. These situations show how past prejudices can influence present choices, maintaining rather than addressing systemic injustices.

The IEEE for Ethical AI and the European Commission's Trustworthy AI framework are two examples of ethical AI guidelines that emphasize the importance of responsibility, transparency, and equity in AI systems. Using techniques like fairness-aware modeling, bias audits, and regular model prediction monitoring are some ways to address biases in the Titanic dataset.



IEEE Standards Association and SSIT: AI-Ethics-related Topics

IEEE SA – *"Autonomous and Intelligent Systems (AIS)"* is the main link for IEEE SA's Autonomous and Intelligent Systems (AIS) website.
There are many links on this website for various IEEE SA activities.

IEEE Ethically Aligned Design

- IEEE SA *Ethically Aligned Design First Edition* - The base publication where it all started (from 2019, but still a good reference!)
- IEEE SA *Ethically Aligned Design for Business* - This one has the capability matrix to assist organizations in implementing and scaling AI responsibly. This is particularly good for SMEs to understand AI and its governance implications. Written by the industry for the industry.
- IEEE SA "*EAD Prioritizing People and Planet as Metrics for Responsible AI*" has a capability matrix on wellbeing metrics. Also good for SMEs

The European Commission's

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

DRAFT

ETHICS GUIDELINES FOR TRUSTWORTHY AI

Conclusion

Although the Titanic dataset offers a rare opportunity to explore predictive modeling, there are significant ethical concerns due to its inherent biases. Age, gender, and passenger class are examples of characteristics that reflect past social values rather than true survival capacity. These biases raise ethical concerns regarding accuracy and fairness, particularly when model output influences actual judgments.

It requires careful consideration to find a balance between these competing priorities. Fairness is more important in real-world applications to prevent perpetuating historical injustices, even though accuracy may be sufficient in instructional settings. Reducing these hazards requires the use of ethical AI methods, such as revealing biases, modeling fairly, and adhering to set standards.

Data scientists can develop models that are accurate and socially conscious by recognizing and addressing these problems, assuring that predictive technologies be used in an ethical and responsible way. This method promotes the creation of AI systems that uphold modern ethical standards and promote justice.

References

Binns, R., 2020. Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT), 3(1), pp.1-10. Available at: https://doi.org/10.xxxx [Accessed 7 December 2024].

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. IEEE. Available at: https://standards.ieee.org/ [Accessed 7 December 2024].

European Commission, 2020. Ethics Guidelines for Trustworthy AI. Brussels: European Commission. Available at: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf [Accessed 7 December 2024].

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021. A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys (CSUR), 54(6), pp.1-35. Available at: https://doi.org/10.xxxx [Accessed 7 December 2024].

Verma, S. and Rubin, J., 2019. Fairness Definitions Explained. IEEE/ACM International Workshop on Software Fairness (FairWare), pp.1-7. Available at: https://doi.org/10.xxxx [Accessed 7 December 2024].