

**ACCURACY OF MACHINE LEARNING ALGORITHMS IN PREDICTING COLLEGE
BASKETBALL GAMES**

ETHICAL IMPACTS OF REGULATIONS ON MACHINE LEARNING

An Undergraduate Thesis Portfolio
Presented to the Faculty of the
School of Engineering and Applied Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Sindhura N. Mente

May 9, 2022

SOCIOTECHNICAL SYNTHESIS

With technological advancements happening more rapidly than ever, it has led to the advent of technologies in Artificial Intelligence (AI) and its subfield Machine Learning (ML). The technical project studies the accuracy of Machine Learning algorithms in predicting college basketball games. The reason for this is both due to the increasing prevalence of the technology and the need for better algorithms and models in the field. The sociotechnical paper aims to examine the current and proposed regulations on Artificial intelligence and determine their adequacy and how they may be giving rise to ethical problems in the field. The project additionally attempts to make suggestions as to where additional regulations should be imposed as a result of the synthesis of the research done. The reason for undertaking the project is that Artificial Intelligence is becoming increasingly integrated into society, and the number of ethical issues surrounding it has also been growing, so better understanding these issues and how to mitigate them is beneficial. The technical paper and the sociotechnical paper are tightly coupled as ML algorithms are a type of AI and their increasing prevalence has given rise to ethical issues which may be exacerbated by the regulations on the technology, which is what the STS project studies.

The technical project involves the building and training of a Machine Learning model to predict the outcome of college basketball games, as the algorithms that currently perform this task are highly biased, which causes skewed results in favor of one team over the other, and have a low threshold of accuracy. The model was built by first obtaining and cleaning the data, building a variety of models to determine the one best suited for the problem, then removing the bias in the data through feature analysis, before finally training the best model for prediction.

After building the preliminary models, it was found that each of them yielded alarmingly low errors of prediction, which was indicative of overfitting of the data. After an analysis of the importance of the features, the ones determined to have caused the overfitting were removed, and the models were rerun, which yielded larger and more reasonable errors that would allow the models to be extrapolated for future use. It was found that logistic regression was able to predict the outcome of a game between two teams with most accuracy. After training and tuning this model, it was used to predict the outcomes of the games in the 2022 March Madness tournament and did so with an accuracy of roughly 62.5%. This shows that it is difficult to predict college basketball games since the rosters change from year to year, so more factors may have to be considered.

The question answered by the sociotechnical project is: How do the regulations, or lack thereof, on artificially intelligent systems affect the ethical concerns surrounding them? The paper discusses how the current and proposed regulations on AI are insufficient and proposes areas to implement them. Research was conducted to understand what the current and proposed regulations, or lack thereof, are and how they are allowing for the issues to increase.

After conducting preliminary research on AI ethics, it was found that data privacy issues, involving corporations including Facebook and Amazon, have been increasing and leading to an increasing collection of user data without consent. there are few loose regulations on AI, which are ineffective. The regulations currently in place include the Trump administration's guidelines encouraging companies manufacturing cars with autonomous driving capabilities, to submit safety self-assessments. However, they are purely voluntary, indicating they are not a truly regulatory measure. The paper concludes by suggesting regulation of AI in the areas of AI safety, approved use, and military applications.

The results of the technical and STS projects are indicative of the integration of AI and ML into daily life, and that this will only increase with time. While there is a push to innovate and develop better models, it should be noted that ethical issues will continue to arise if not addressed. Therefore, there must be an equal emphasis on both the regulation of AI and innovations in AI.

TABLE OF CONTENTS

SOCIOTECHNICAL SYNTHESIS

ACCURACY OF MACHINE LEARNING ALGORITHMS IN PREDICTING COLLEGE BASKETBALL GAMES

with John Kim

Technical advisor: Haiying Shen, Department of Computer Science

ETHICAL IMPACTS OF REGULATIONS ON MACHINE LEARNING

STS advisor: Catherine D. Baritaud, Department of Engineering and Society

PROSPECTUS

Technical advisor: Haiying Shen, Department of Computer Science;

STS advisor: Catherine D. Baritaud, Department of Engineering and Society