Feasibility of Utilizing Machine Learning to Identify a More Sustainable Alternative to Polyester in Textiles

Shriti Dham
Northgate High School, California, 94598

Abstract

The mass production of textiles has created several concerns for the environment and its future, with the increase of greenhouse gases that are released during the production of Polyester. Machine learning is a recent technology commonly used in polymer science to consider vast amounts of data, relatively quickly. The aim of the project is to find an alternative polymer that possesses similar characteristics to polyester with a considerably lower environmental impact. In this study, we used a publicly available dataset, *PI1M*, containing information on over one million polymers to train artificial neural net. We looked into predicting key features for these polymers, including glass transition temperature, density, melting temperature, oxygen permeability, compressibility, and bulk modulus. Through the process of machine learning modeling, several properties of new polymers were successfully predicted. An agreement was found between the true and predicted values for each of these features. The results show the presence of a possibly more sustainable alternative to Polyester in textiles. This study demonstrates the feasibility of using machine learning to discover new, more sustainable polymers.

Introduction

Annually, over 92 million tonnes of textiles are discarded around the globe [1]. This presents major consequences for the environment and has already contributed to extensive pollution. A substantial amount of these textiles consist of fibers with a polyester base, also known as Polyethylene terephthalate or, more commonly, PET [2] . Polyester is the most produced textile–over 50 billion tonnes–in the world, largely due to its multitude of desirable properties in clothing. By itself it has benefits, but especially when compounded with other materials, the outcome possesses extended useful qualities. The matter is quite durable compared to competing clothing materials and blends with other components smoothly. As well, the material can be elastic and flexible, making it preferable for clothing. It retains dye relatively easily and is hydrophobic, which makes it imperviable to common outdoor conditions [3]. Above all, producing Polyester is extremely cost effective, being half the price of one kilo of cotton [4]. Although the cost appeals to both manufacturers and consumers, leading to increased sales, it also creates an equal amount of waste of the product, which has harshly affected the environment for many years, and continues to pose threats (see Figure 1).

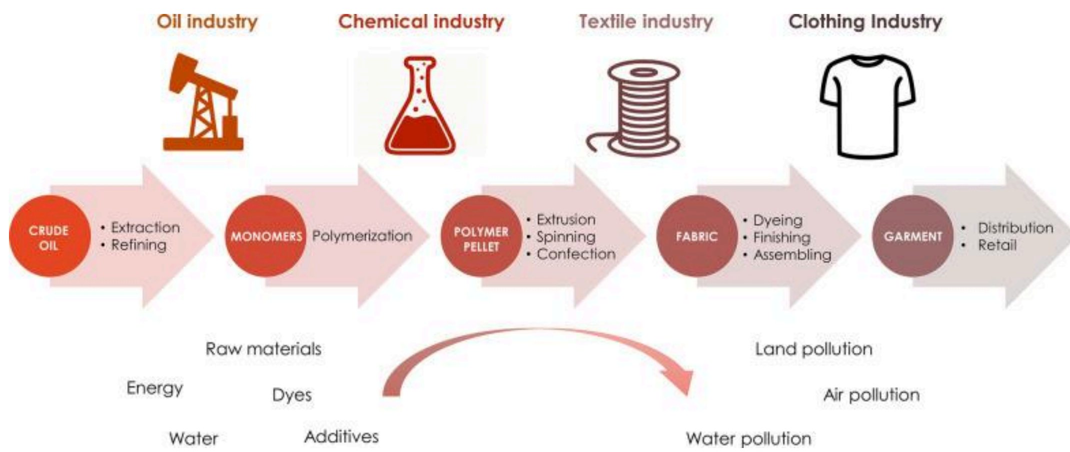

Figure 1. Pollution from Manufacturing [2]

Textiles are most commonly disposed of by methods of incineration or landfill which both pose significant threats to the ecosystem and atmosphere. Less than 1% of PET clothing is recycled into new clothing [5]. Polyester is a form of plastic, a manufactured, synthetic material. Plastics are notorious for their persistence to degradation. This specific type of plastic can remain in ecosystems for up to two hundred years [6]. The other method, incineration, poses risks of its own. Incineration is the burning of products that emit greenhouse gasses and toxins into the atmosphere, which damages the ozone layer and heightens global temperatures. This presents the immediate need for a replacement of Polyester that leaves a less substantial impact, especially when disposed of in landfills and incineration plants.

Some alternatives to polyester currently exist. Recycled polyester is a sustainable alternative made from plastic waste, and significantly reduces total energy consumption. Still, the presence of polyester, even in the recycled version, causes microplastics in the environment. Another material to substitute polyester that is currently being used is biodegradable polyester, made from natural biomass. This helps to reduce waste in landfills and pollution of the land, but producing it and even while it degrades, it can release greenhouse gasses.

There are thousands of different polymers that can be considered, some are synthesized by humans in labs, while others are naturally occurring. Many are available online, compiled into large datasets. The aim is to evaluate a large number of polymers for replacing polyester, which demonstrates why machine learning is best fitting to reach the goal. The machine learning model has the ability to grasp these immense amounts of data. The type of machine learning model used is called a neural network, which identifies patterns in the set to create a specialized output [7] (see Figure 2). After analyzing the data, it can make predictions, which is essential to the outcome of this research. The prediction of new polymers is necessary, as existing ones similar to Polyester are extremely harmful to Earth's ecosystems.

Models have been trained in the past to predict polymers and have been successful in the task. Datasets like *PolyInfo* were used in machine learning models to predict polymers. In the article "Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm" it states that, "Using a molecular design algorithm trained to recognize quantitative structure—property relationships with respect to thermal conductivity and other targeted polymeric properties, we identified thousands of promising hypothetical polymers [13]. This is similar to the task that the machine learning model in this research performed.

Methods

In order to train the machine learning model, it was necessary to gather a multitude of databases that provided us with computational and experimental properties of polymer materials. These databases were found through academic literature accessed through online libraries and search engines, specifically Google Scholar. We researched various databases of polymers based on a specific property that the polymers possessed. These databases consisted of the polymer in the form of a SMILES string along with a numerical value of the property. A SMILES string refers to a Simplified Molecular Input Line Entry System. This is a notation that represents the chemical structure of a certain polymer, with certain symbols representing atoms and bonds [8]. It is especially useful in computer programs, and is required to identify the specific polymers that can replace Polyester. The properties we primarily focused on were density, glass transition temperature, oxygen permeability, melting temperature, compressibility, and bulk modulus. These characteristics are important to the performance of the polymer as a textile, and can have effects on its sustainability index. For each specific property, we trained a model to predict values of it for different polymer materials.

Firstly, we had to make sure to fit the parameters of the neural network model in order to use the data. The training process allows for this to occur. To train the model and predict properties of polymers, we used the common coding language *Python* in the software package by the name of *Chemprop*, which is described as "a repository containing message passing neural networks for molecular property prediction" [9]. By default, *Chemprop's* system used 80% of the given data to train the neural network (see Figure 2). As for the other twenty percent, ten is used for validation of the model and the last ten is used to test. The twenty percent used at the end contributes to the accuracy of the predictions. As well, it helps to identify errors the model may have made.

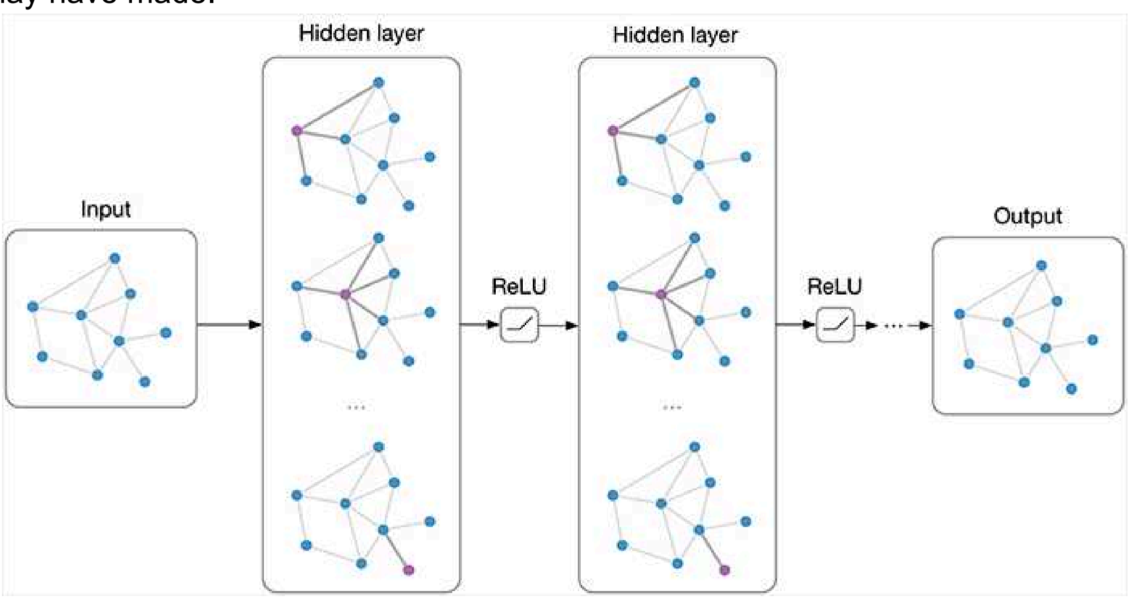

Figure 2. Graph Neural Network Model [14]

After acquiring the predicted values, we started the screening process. We used the previously trained models on the smaller property databases to predict on a large 1 million polymer database. This database, *PI1M*, includes publicly available information. *PI1M* is a benchmark database containing about one million polymers for polymer informatics, made specifically for machine learning research [10]. It was previously trained on twelve thousand polymers, which gave it the data to generate its own polymers. The polymers from this were used as a prediction dataset for creating the new polymers.

Using search engines, I obtained the values of the density, the glass transition temperature, the oxygen permeability, the melting temperature, compressibility, and the bulk modulus of Polyester, being the target polymer. It was then necessary to convert the researched value into the correct units used in the values inside the databases. Once converted, we plugged in the actual Polyester values into the code. We screened the large *PI1M* database for polymers that most closely matched the target values for polyester.

Results

Using the application *Visual Studio Code*, we were able to compile the predicted values of each property into its own scatter plot. Each plot displayed a regression line, and all took on a linear nature. The input value on the x-axis of the plot represented the predicted value from the model and the output value on the y-axis of the plot was the true value of the property (see Figures 3-8).

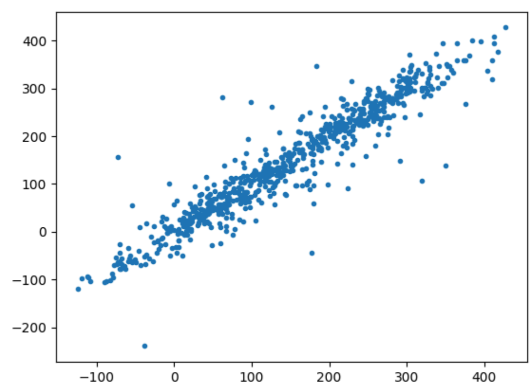

Figure 3. Glass Transition Temperature True vs. Predicted

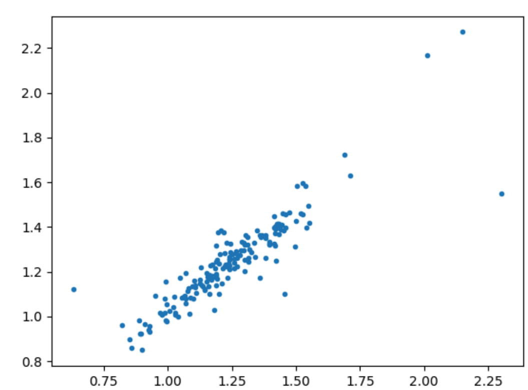

Figure 4. Density True vs. Predicted

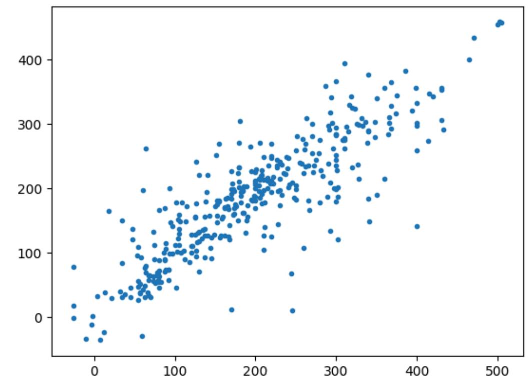

Figure 5. Melting Temperature True vs. Predicted

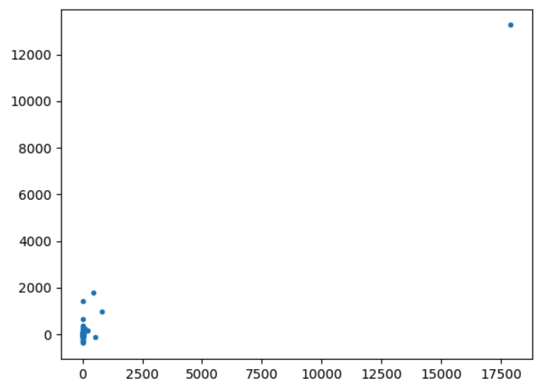

Figure 6. Oxygen Permeability True vs. Predicted

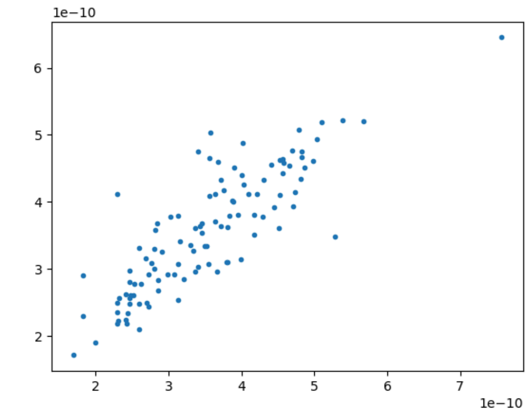

Figure 7. Compressibility True vs. Predicted

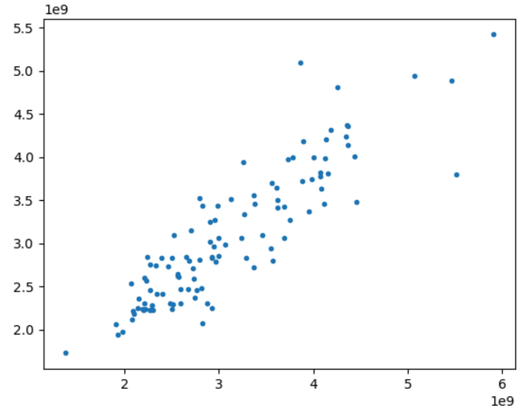

Figure 8. Bulk Modulus True vs. Predicted

As well, the code revealed the mean absolute error, which was higher for some property datasets than others. The mean absolute error is used to evaluate models. It is a measure of errors in predicted values, or regression problems [11]. More specifically, it is the mean of the absolute difference of the true and predicted values. The lower the mean absolute error, the more accurate the findings and predictions. Each of the property databases varied in size, some consisting of less than a thousand polymers, and others consisting of hundreds of thousands, which affected the value of error and how the regression line looked. This is apparent in the oxygen permeability scatter plot. The plot contains considerably less points than the other graphs, as the dataset has much less values. It affected the concentration of points, but a linear pattern is still somewhat recognizable.

The *python* script allows us to enter a numerical value of polymers in the database to focus on. Based on the amount of polymers entered, the model will find the closest polymers in value to the values of Polyester. To illustrate, if the number 50000 is entered, the model will screen the database for the closest fifty thousand polymers in value. The *Python* script that displayed the polymers with the most similar values to the target set up the results of the experiment. The polymers closest to the target values were separated by property, and by number of polymers in the file. To be able to get a wide variety of results, we screened the database for the top fifty thousand, the top hundred thousand, the top one hundred and fifty thousand, and lastly the top two hundred thousand closest polymers. The smaller the number of polymers one looks for, the fewer properties will match with Polyester's. The amount of polymers the database can screen will lessen, therefore the chance of a matching polymer will lessen. However, it is also advantageous because the values will be closer to the target. Based on the collected data the most amount of properties that paralleled with Polyester's was when screening for the top two hundred thousand closest polymers. The results satisfied all six of the properties tested, showing forty polymers would fall within the top twenty percent of the closest polymers in value. For one hundred fifty thousand polymers, or the closest fifteen percent of values, the model found seven polymers that would match closely to Polyester. The polymer from this screening that's values matched most closely was "*OC(c1ccccc1)C(*)=O" (see Table 1).

Table 1. Property Values of Polyester vs. *OC(c1ccccc1)C(*)=O

	Density (g/cm^3)	Melting Temperature (°C)	Oxygen Permeability (Barrer)	Compressibility (Pa)	Bulk Modulus (Pa)	Glass Transition Temperature (°C)
Polyester	1.25 g/cm^3	295°C	0.0457 Barrer	(4e-10) Pa	2.5e9 Pa	68.3333 °C
*OC(c1ccccc1)C(*)=O	1.2451408 g/cm^3	265.16592 °C	38.90782 Barrer	4.165523e-10 Pa	2597991000.0 Pa	81.75616 °C
Residual	0.00486	29.83408	-38.86212	-1.65523e-11	-97991000	-13.42286
Percent Difference	0.3888%	10.1132%	8.5037e4%	4.1381%	3.9196%	19.6432%

Additionally, we tested the top ten percent of the dataset, or one hundred thousand polymers. The model revealed that twenty polymers would fall within the closest ten percent of the database's values, for five out of six of the properties. The properties it would satisfy are the density, the melting temperature, the oxygen permeability, the compressibility, and the bulk modulus value. The one property that did not fall within the top ten percent was the glass transition temperature, which is quite valuable in clothing, as it can consider the flexibility, heat exposure, or wearing conditions the fabric can take.

Conclusion and Future Work

The target of this research was to identify if machine learning models could be harnessed to predict and create new polymers that can be used as a more sustainable alternative to Polyester, specifically in textiles. It was necessary to train the machine learning models with a database of pre-existing polymers, so it could then generate predictions of new polymers. We screened a large database of 1 million polymers for matches to property values of Polyester. The process created novel polymers which were somewhat similar to Polyester, except for environmental impact. This can expose what alternatives come out of the experiment and assist in determining whether or not Polyester can be substituted for another compound in textiles, without giving up its features.

A future course of action would be to identify which material properties correlate with environmental metrics. This is necessary to determine if the new, predicted polymer would be sustainable.

We considered utilizing a self supervised pre training script for this code. Using a self supervised pre training script can improve accuracy, as it can learn from unlabeled data. This will help with improving the machine learning model's performance, especially when asked to carry out a specific task, in this case the task was finding the most similar value between two datasets [12].

Synthesis of material and property testing can be done to verify the performance of the polymer as both a textile and environmentally friendly material. The carbon footprint of the polymer and its sustainability would need to be considered, more seriously than its performance. If the polymer would perform well in each necessary area, it can be seriously considered as a working alternative to Polyester in clothing textiles.

References

[1] "Unsustainable fashion and textiles in focus for International Day of Zero Waste 2025." *UNEP*, 27 March 2025, https://www.unep.org/news-and-stories/press-release/unsustainable-fashion-and-textiles-focus-international-day-zero. Accessed 4 December 2024.

[2] Palacios-Mateo, Cristina et al. "Analysis of the polyester clothing value chain to identify key intervention points for sustainability." *Environmental sciences Europe* vol. 33,1 (2021): 2. doi:10.1186/s12302-020-00447-x

[3] Jaffe, Michel, et al. "Polyester fibers." *Thermal Analysis of Textiles and Fibers*, 2020, pp. 133-149. *Science Direct*, https://www.sciencedirect.com/science/article/abs/pii/B9780081005729000082. Accessed 2 November 2024.

[4] "Fossil fashion: the hidden reliance of fast fashion on fossil fuels • Changing Markets." *Changing Markets Foundation*, https://changingmarkets.org/report/fossil-fashion-the-hidden-reliance-of-fast-fashion-on-fossil-fuels/. Accessed 2 November 2024.

[5] Montoya, Juliana. "The fashion industry needs an alternative to recycled PET." *Plastics Engineering*, 14 November 2023, https://www.plasticsengineering.org/2023/11/the-fashion-industry-needs-an-alternative-to-recycled-pet-002381/. Accessed 18 June 2025.

[6] Stanes, Elyse, and Chris Gibson. "Materials that linger: An embodied geography of polyester clothes." *Geoforum*, vol. 85, no. October, 2017, pp. 27-36. *Science Direct*, https://www.sciencedirect.com/science/article/abs/pii/S0016718517301835?via%3Dihub. Accessed 2 November 2024.

[7] Hardesty, Larry. "Explained: Neural networks." *MIT News*, 14 April 2024, https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414. Accessed 4 December 2024.

[8] Öztürk, Hakime, et al. "A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction." *BMC Bioinformatics volume*, vol. 17, no. 128, 2016. *Springer Nature*, https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-0977-x#:~:text=SMILES%20strings%20convey%20information%20about%20molecular%20structures,atoms%20and%20=%2C%20%23%20for%20bonds%20(www.daylight.com/dayhtml/doc/theory/theory.smiles.html. Accessed 4 December 2024.

[9] Heid, Esther, et al. "Chemprop: A Machine Learning Package for Chemical Property Prediction." *Journal of Chemical Information and Modeling*, vol. 64, no. 1, 2024, pp. 9-17. *Git Hub*, https://doi.org/10.1021/acs.jcim.3c01250. Accessed 16 September 2024.

[10] Xu, Pengcheng, et al. "New Opportunity: Machine Learning for Polymer Materials Design and Discovery." *Advanced Theory and Simulations*, vol. 5, 2022. Accessed July 2024.

[11] Hodson, Timothy O. "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not." *European Geosciences Union*, vol. 15, no. 14, 2022, pp. 5481-5487. *Geoscientific Model Development*, https://gmd.copernicus.org/articles/15/5481/2022/gmd-15-5481-2022-discussion.html. Accessed 4 December 2024.

[12] Tayebi Arasteh, Soroosh, et al. “Enhancing diagnostic deep learning via self-supervised pretraining on large-scale, unlabeled non-medical images.” *European Radiology Experimental*, vol. 8, no. 10, 2024. *Springer Open*, https://eurradiolexp.springeropen.com/articles/10.1186/s41747-023-00411-3#:~:text=SSL%20is%20a%20learning%20paradigm,2021%20by%20Dosovitskiy%20et%20al. Accessed 4 December 2024.

[13] Wu, Stephen, et al. “Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm.” *npj Computational Materials*, vol. 5, no. 66, 2019, https://www.nature.com/articles/s41524-019-0203-2. Accessed 29 December 2024.

[14] Karagiannakis, Sergios. “Graph Neural Networks - An overview.” *AI Summer*. 1 February 2020, https://theaisummer.com/Graph_Neural_Networks/. Accessed 22 June 2025.