UNIVERSITY OF TURKU

# Predicting Tire Compound Properties Using Machine Learning

Materials Engineering

Master's thesis

Author:

Anna Salmikangas

9.10.2025

Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

**Subject**: Materials Engineering
**Author(s)**: Anna Salmikangas
**Title**: Predicting Tire Compound Properties Using Machine Learning
**Supervisor(s)**: PhD Tomasz Galica, DI Marko Ylitolva
**Number of pages**: 60+3 pages
**Date**: 9.10.2025

Safety, performance, and cost are key requirements when developing new tire compounds. In addition, environmental questions related to tire materials are increasingly rising which introduces a lot of requirements for materials used in tires. Compound development takes a lot of time and money as multiple laboratory tests are needed in order to make sure the compound meets the requirements. The purpose of this thesis is to study if machine learning methods could be used with real life experimental data to predict the properties measured in laboratory. Common compound materials and laboratory testing methods are discussed in the theory part.

The experimental part consists of comparison between six different machine learning methods: linear regression, ridge regression, kernel ridge regression, support vector regression, random forest regression, and gradient boosting regression. Models were created and evaluated using leave-one-out cross-validation (LOOCV) and leave-one-group-out cross-validation (LOGOCV). Evaluation metrics were $R^2$, normalized RMSE, and normalized MAE to overcome the different scales and units of the properties. Kendall's $\tau$ was used as dimensionless metric to evaluate how well the models can maintain the ranking in the data.

The data was combined from 9 laboratory experiments with the total of 86 samples. The aim was to predict eleven properties that are commonly used in the tire industry based on the compound material amounts. Studied properties are maximum torque, minimum torque, difference of maximum and minimum torque, cure time, tensile strength, modulus 300, elongation at break %, hardness, tan$\delta$ (0°C), tan$\delta$ (60°C), and E' (30°C).

There was no single all-round best model and there were big differences in models' prediction performance for different properties. On average, kernel ridge regression showed promising results when using LOOCV. The metrics showed poor performance for LOGOCV. This indicates that the machine learning models can be used to predict properties within modelled material distribution but extrapolating models to new materials possess significant challenges.

**Key words**: machine learning, tire compounds, rubber compound, compound properties

## Acknowledgement

This thesis was done as a commission to Black Donuts Inc. I would like to express my gratitude for Black Donuts Inc. for commissioning this thesis and providing me the opportunity to work on this topic. Their support and expertise made this thesis possible. I also want to thank my academic supervisor Tomasz Galica for guiding me through the process. A special thanks to my husband for always believing in me.

# Table of contents

# Abbreviations

| | |
|---|---|
| ANN | Artificial Neural Network |
| BR | Polybutadiene Rubber |
| CV | Coefficient of Variation |
| DMA | Dynamic Mechanical Analysis |
| EB % | Elongation at Break % |
| GBR | Gradient Boosting Regression |
| GPR | Gaussian Process Regression |
| IID | Independent and Identically Distributed |
| KRR | Kernel Ridge Regression |
| LOGOCV | Leave-One-Group-Out Cross-Validation |
| LOOCV | Leave-One-Out Cross-Validation |
| LR | Linear Regression |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MDR | Moving Die Rheometer |
| $M_H$ | Maximum Torque |
| $M_H$-$M_L$ | Difference between Maximum and Minimum Torque |
| $M_L$ | Minimum Torque |
| ML | Machine Learning |
| NR | Natural Rubber |
| PHR | Parts per Hundred Rubber |
| RF | Random Forest |
| RMSE | Root Mean Squared Error |
| RPA | Rubber Process Analyzer |
| RR | Ridge Regression |
| RSM | Response Surface Method |
| RSS | Sum of Squared Residuals |
| SBR | Styrene-Butadiene Rubber |
| SVR | Support Vector Regressor |
| $t_{90}$ | Cure Time |
| XGB | Extreme Gradient Boosting |

# Symbols

| | | | |
|---|---|---|---|
| $\alpha$ | Lagrange multiplier | $\Phi$ | Mapping from the input space to feature space |
| B | Number of trees in tree models | P | Number of concordant pairs |
| $\beta$ | Coefficient factor for a variable | p | p-value of statistical significance |
| $\hat{\beta}$ | Estimated coefficient | Q | Number of discordant pairs |
| C | Regularization parameter in SVR | $r_i$ | Residual of $i^{th}$ term |
| d | Thickness | $\sigma$ | Standard deviation of the training samples |
| $\delta$ | Phase angle | T | Number of tied pairs in x |
| E' | Storage normal modulus | $\tau$ | Kendall's correlation coefficient |
| E'' | Loss normal modulus | Ts | Tear strength |
| E* | Complex normal modulus | $\Theta$ | Independent random vector |
| $\varepsilon$ | Tolerance margin in SVR | U | Number of tied pairs in y |
| $\epsilon_i$ | Error of the $i^{th}$ term | w | Coefficient vector |
| F | force | x | Training samples |
| $f$ | Function | $x_i$ | Independent variable $i^{th}$ term |
| $\mathcal{F}$ | Reproducing Kernel Hilbert Space (RKHS) | X | Feature matrix |
| $\hat{f}$ | Predicted function | y | Dependent variable |
| $\gamma$ | Kernel coefficient | $\hat{y}$ | Predicted y |
| $\infty$ | Infinity | $\hat{y}_i$ | Predicted $i^{th}$ term |
| k | Number of features | $y_i$ | Actual $i^{th}$ term |
| $\lambda$ | Shrinkage parameter | $\bar{y}_i$ | Mean of the observed data |
| $\mu$ | Mean of the column | z | Scaled value |
| N | Number of data points | | |

# 1 Introduction

Tires are a crucial part of modern transportation on land as they are usually the only contact point between the vehicle and the road. Tires impact vehicle safety and performance to a great extent. In tire development, the balance is sought between the highest grip (safety), lowest rolling resistance (fuel economy), and wear resistance (long product lifetime). These properties are affected by the materials, tread pattern, design, and manufacturing process [1]. Ideal tires are safe and easy to handle with low energy consumption and long lifetime. Tires play important role in environmental questions, as they are one of the biggest contributors to microplastic in the environment from tire, road, and brake wear [2]. Climate change and efforts against environmental pollution require tires to be made of more environmental-friendly and less polluting materials.

Each tire has a specific rubber formulation that affects its price, processing, and properties. These three factors are the main influencers of selecting the most suitable material blend. In industry, these material blends are called recipes, and they describe the proportions of the materials of which tires are made. [3] There are hundreds of different materials that can be used in the recipe with unlimited number of combinations, and the suitability of the recipe is tested in laboratory which is both expensive and time-consuming.

Tire recipe development has been based heavily on trial and error as well as experience. Regardless, many tire manufacturers have announced R&D projects related to machine learning and artificial intelligence in recent years [4]. Different machine learning methods have been used to predict properties of rubber materials [5,6]. Machine learning can help the tire development process by predicting material properties before laboratory tests, so that the recipe can be optimized before any testing and then verified in the laboratory to avoid unnecessary testing rounds, and the most suitable recipe can be found faster.

This thesis was commissioned by Black Donuts Inc. to study machine learning in tire material development. The aim is to answer following questions: Can machine learning methods predict tire material properties based on their formulation alone? Which machine learning method performs best for this prediction? Do some properties have higher prediction precision?

The rest of the thesis is organized as follows. Chapter 2 briefly describes tire materials and laboratory testing. Chapter 3 reviews machine learning methods suitable for predicting the

target properties in this thesis. In chapter 4, methodology is introduced for the study. The results are presented in chapter 5. Chapter 6 discusses the results and their significance, and chapter 7 contains the conclusions of this thesis.

# 2 Tire compound materials

Tires can be found widely in transportation sector. Although they seem to a be simple, black torus, tire development is a complex process influenced by safety, performance, and cost requirements. There are many parameters that are involved in tire performance with countless different parameter combinations. Changing one parameter will often influence others. [7] When it comes to developing new tire materials, there are three important factors: price, processability, and properties [3]. The combination of these aspects defines what tire materials end up in production, and the designer must find the right balance between these three.

## 2.1 Tire compounds

Rubber is the main component for tires, but it is not the only one as rubber alone does not make very good tires. Process of modifying a rubber or a rubber blend is called compounding, where suitable ingredients and their quantities are selected. The resulting mixture is called compound. [3] Compounding optimizes properties and aims to meet all the requirements set for a tire.

The final properties depend on the type of rubber(s) and the additives used. The properties can be further modified by changing the processing steps or conditions. Variables such as speed, pressure, and temperature can be changed during mixing, extrusion, calendaring, and molding [7]. Finally, both the raw materials and the processing affect the final cost of the tire and decide whether it is profitable to produce it.

Compound recipes form the basis of tire manufacturing. Usually, a blend of two or more different elastomers forms the base of the recipe, and all the other ingredients are given against rubber. The proportions of different materials are expressed as PHR (parts per hundred rubber). This enables easier comparison between different compounds. [3] Typically, a tire consists of approximately ten different ingredients, but the amount varies based on the application and there are hundreds of commonly used raw materials to choose from with different grades. Even having a different supplier can change the properties of the raw material.

Tire compositions are not uniform, and different materials are used in different parts of a tire. A tire tread is in contact with the road and it consists of rubber typically compounded with

carbon black, silica, oils, and vulcanizing chemicals. In the other parts of the tire, steel, textile, and other chemicals are used to get the desired properties. [7]

In the following section of this chapter, typical materials for tire tread are introduced and discussed.

### 2.1.1 Elastomers

Rubber is the main component of tire and accounts for 40-50 % of the tire compound [8]. Mostly used rubber types in tires are natural rubber (NR), styrene-butadiene rubber (SBR) and polybutadiene rubber (BR). NR, SBR, and BR are unsaturated elastic polymers. Unsaturation gives them elastic properties but also make them more vulnerable to aging due to environmental factors [9]. Each of these rubbers have different strengths and weaknesses, and they are often mixed to balance the compound. Rubbers are used because of their elastomeric nature that provides grip on various surface and flexibility while maintaining its shape. Rubber can hold the weight of the load and make driving more comfortable with cushioning. [7] Cost is another factor, as rubber is easily available for relatively cheap price as natural and synthetic rubber even if the price has been rising during the recent years [10].

NR is obtained from rubber trees by coagulation of latex. It has very high elasticity, good tensile strength, and good abrasion resistance but it degrades easily when aged and exposed to oils. Originally, tires used NR, but nowadays it is often compounded with other elastomers to resist aging. [9] The proportion of NR is higher in larger tires for commercial and off the road applications as it provides lower operating temperature, reduced rolling resistance, tear strength and good adhesion for components to increase durability and tire retreadability [8].

SBR is the copolymer of styrene and butadiene, and it is the most common synthetic rubber in tires. It is often mixed with NR in tires. SBR has a good abrasion, and it ages better compared to NR. [9] Nowadays it is added for wet slide and traction properties with good abrasion resistance [11]. BR increases resistance to abrasion and fatigue but also for cut propagation [11]. BR has low glass transition temperature which makes it very resistant to abrasion and gives low rolling resistance, but it also leads to poor wet traction. The processability of BR is weaker compared to NR and SBR for which it is always used in blends. [9]

### 2.1.2 Fillers

Fillers are the second biggest component in tire compound, commonly 30 % or more of the total weight [1]. They are used as reinforcement aids to make the tires last longer in use and for cost reduction. Typical fillers are carbon black and silica, but there are other options like chalk and clay.

Carbon black is the most commonly used filler, and it is usually the second raw material in weight after the rubber. There is a wide range of carbon blacks available which makes the selection crucial. The most important properties in carbon black selection are particle size, particle surface area, particle surface activity, and particle shape which all affect how well the carbon black works with the rubber matrix. Carbon black improves the tensile strength, tear strength, and abrasion resistance together with the hardness and modulus while elongation at break and volume swell in fluids decreases. [9]

Silica is used to improve tear strength and reduce heat buildup and rolling resistance. Silica is a polar material and it is not naturally compatible with non-polar elastomers so it is usually used with coupling agents. [12] Compared to carbon black, silica gives lower rolling resistance, and improvements to trouser tear strength, heat resistance, and adhesion to fabric and metal. The drawbacks are high viscosity, stiffness, and hardness (compared to carbon black). [13]

### 2.1.3 Sulfur Vulcanization

Vulcanization (also known as curing) is important process step to create crosslinking between the elastomers. Vulcanized elastomer is stronger, less sticky, and retracts better to their original shape after deforming force [14]. Compared to unvulcanised rubber, vulcanization gives many useful mechanical properties such as resistance against swelling in liquids, temperature changes, and light [15].

Sulfur is the most used vulcanization agent. It reacts with the unsaturated groups of NR, SBR, and BR and forms crosslinked network between the molecules [9]. Sulfur is preferred for its economics and the controllability it gives together with accelerators to vulcanization rate and crosslinking nature [12].

Sulfur is used together with accelerators and activators. The use of accelerators and activators shortens the vulcanization time significantly from hours to 1-3 minutes [14]. Less sulfur is

needed when the crosslinking is more efficient and unwanted side reactions can be avoided. Using less sulfur also improves the aging properties of the vulcanized rubber. Accelerators also permit lower vulcanization temperature which prevents heat damage to rubber [12]. Accelerators are categorized to primary and secondary accelerators. Primary accelerators provide scorch delay, medium to fast cure, and good modulus development whereas secondary accelerators produce fast curing stocks. Common primary accelerators are thiazoles MBT, MBTS, and sulfenamides CBC, TBBS, MBS, and DCBS. Typical secondary accelerators are guanides DPG, DOTG, thiurans TMTD, TMTM, and dithiocarbamates ZMDC, and ZBPD. [15]

Activators are used to get the full potential of the accelerators. Without accelerators, activators do not increase the cross-linking [15]. Used activators are different inorganic and organic chemicals, most commonly a combination of zinc oxide and stearic acid which are competitive in terms of economic and ecological reasons. Accelerators react with zinc oxide and stearic acid and form a complex that reacts with the sulfur which in turn reacts with the polymer. Addition of zinc oxide increases the crosslinking density and leads to high tensile strength and stress values. Meanwhile, stearic acid improves processability and distribution of the filler. [12]

### 2.1.4 Plasticizers and processing aids

Plasticizers improve the processability of the compound, dispersion of fillers in the compound, and influence properties. Commonly used plasticizers are petroleum oils, pine tar, resins and waxes. [1] Plasticizers can be divided into processing oils and extender oils. The former ones improve the flow properties and processability whereas latter ones are used to extend and lower the price of the product. Primary plasticizers decrease the viscosity of the unvulcanized compound by dissolving or swelling the polymer which leads to lower hardness and better elasticity in the vulcanized product. Secondary plasticizers do not dissolve the polymer but act as a lubricant in the compound improving moldability. [12]

Mineral oils are cheap and thus used in large quantities in the tire industry. Low dosages are used for processing aids, and high dosages for extending. Tires typically use aromatic mineral oils. Mineral oils are added to some synthetic rubbers, such as SBR, already during production of the rubber. These rubbers are called oil-extended polymers. More expensive synthetic oils are used to get special properties for the rubber. [12]

Processing aids are also used to improve processability but without plasticization. They are usually used in low dosage levels and have little to no effect on physical properties. Processing aids are typically classified by their function but some of them are multifunctional. Different functions are viscosity reduction, polymer blending, dispersion, lubrication, tackiness, reinforcement, and demolding the cured rubber. Typical process aids are different resins, metallic soaps, and fatty acids. [12]

### 2.1.5 Stabilizers

Rubber starts to crack when aged due to environmental factors. Many additives are used in tires as stabilizers make the tire more resistant to environmental challenges like oxygen, ozone, heat, and UV light. To overcome these challenges, antidegradants are added to the compound.

Antidegradants include antioxidants and antiozonants with typical levels around 0.5-2.5 PHR. Antioxidants protect the compound against aging and oxidation. 2,2,4-Trimethyl-1,2-dihydroquinoline (TMQ) is commonly used antioxidant but it lacks antiozonant effect. It is often paired with the antiozonant N-phenyl-N'-(1,3-dimethylbutyl)-p-phenylenediamine (6-PPD). [12] 6-PPD reacts well with ozone and thus lengthens the product lifetime which makes it very popular in tires. Unfortunately, alternatives are searched as 6-PPD with its transformation products is found to be toxic in environment [16].

### 2.1.6 Summarization

Table 1 summarizes the key properties of different compound components with example materials and a typical PHR range (Based on [9,13]). Ciullo and Hewitt [17] give a reference for a suggested formula of tire base tread. The recipe is presented in table 1.

Table 1 Summary of tire compound materials

| Component | Example materials | PHR range | Properties | Example Recipe (PHR) |
|---|---|---|---|---|
| Elastomer | NR<br>SBR<br>BR | 100 | Resilience<br>Environmental Resistance: Temperature, oil, UV, ozone, solvents<br>Reinforcement<br>Processing requirements<br>Cost | Natural Rubber (60.0)<br>Cis BR 40.0 (40.0) |
| Filler | Carbon Black<br>Silica | Up to 50 | Reinforcement (abrasion, tear)<br>Durometer hardness<br>Processing (viscosity)<br>Colour<br>Cost | N-660 Black (45.0) |
| Crosslinking system | Sulfur<br>Sulfenamides<br>Zinc Oxide<br>Stearic Acid | Up to 15 | Cure state<br>Hardness<br>Modulus<br>Set<br>Heat build-up<br>Scorch safety<br>Heat and aging resistance<br>Cost | Zinc Oxide (5.0)<br>Stearic Acid (1.5)<br>80 % Oiled Insoluble Sulfur (3.25)<br>Delac® NS -TTBS (1.0)<br>DTDM (1.0) |
| Plasticizers | Petroleum Oil<br>Resin<br>Wax | Up to 50 | Processing<br>Viscosity, mixing, extrusion<br>Calendering<br>Hardness<br>Elastomer compatibility<br>Low temperature flexibility<br>High temperature stability<br>Cost | Naphthenic Oil (6.0)<br>Hydrocarbon Resin (2.0) |
| Stabilizers | TMQ<br>6-PPD | 0-5 | Heat resistance<br>Ozone resistance<br>Staining, colour | Naugard® Q – TMQ (0-2.0) |
| Other | Organosilanes<br>Other chemicals | | Adhesion promoter<br>Cell formation<br>Coupling silica to elastomer | |

## 2.2 Laboratory Testing

As described above, rubber compounds have a wide range of unique properties that are not related to each other. These unique properties make testing and describing the compound complex when changing one material or even material grade can lead to different behavior in the compound. The main ingredient, rubber, is different from many other engineering materials. Stress-strain relationship is nonlinear, Hooke's law is only applicable up to 50 % extension, deformation leads to large hysteresis effect, and mechanical behavior is unpredictable even when stiffness and viscosity are known [9]. This results in various tests that are needed to get the wanted properties for a rubber compound.

Compounds need to be tested before they are applied in large-scale manufacturing. Several physical properties are tested in the laboratory to determine if the compound can be used in production. The compound usually needs the finetuning of the selected raw materials to get the wanted balance of the properties. [18] Tires are then manufactured in batches that use the tested compound recipe. The common tests include vulcanization characteristics, tensile stress-strain, hardness, wear resistance, cut and chip characteristics, tear strength, and viscoelastic behavior.

### 2.2.1 Vulcanization characteristics

Vulcanization characteristics (also known as rheometric properties) describe how well the compound is vulcanized. Interesting characteristics are, how long does it take for crosslinking to start, the rate of crosslinking, and the extent of crosslinking [14]. Crosslinking density is proportional to the stiffness of the rubber which is used in the measurements [19]. Vulcanization characteristics of a compound batch are typically measured using MDR (Moving Die Rheometer) that measures the development of the torque in time.

Minimum torque ($M_L$) is the lowest torque at the beginning of the vulcanization test. It tells about the processability of the compound. Too high or too low values often indicate problems in the batch. Maximum torque ($M_H$) is the highest torque value in given time. It tells the final state of the vulcanized rubber, and it indicates the crosslink density. Both $M_L$ and $M_H$ are derived from torque-time curve. Difference of maximum and minimum torque ($M_H$-$M_L$) tells the change in stiffness during vulcanization. Scorch time and cure time are points where 10 % ($t_{10}$) and 90 % ($t_{90}$) of vulcanization is completed. $t_{10}$ is important to know, as premature

curing can ruin the whole batch. $t_{90}$ is considered to be the optimum cure time in the industry to avoid overcuring. [20]

### 2.2.2 Mechanical properties

Mechanical tests include a variety of different tests that are used to evaluate mechanical performance of the compound. Multiple instruments are used in mechanical testing.

Tensile stress-strain is widely used in the rubber industry, and it is one of the oldest tests in the industry. A dumbbell shaped sample piece of cured compound is pulled apart at constant rate until it breaks. Tensile testing can reveal problems or errors in manufacturing such as poor mixing and dispersion, contaminants in compound, or problems in curing. Tensile strength (in MPa) is the maximum stress the sample can stand before breaking. Elongation at break % tells how much the rubber can be stretched before breaking. Tensile stress is often reported at certain elongation e.g. 100 % modulus or 300 % modulus (in MPa) even when they are not actual modulus values. [18]

Tear characteristics are related to compounds crosslink density and state of cure, filler type and loading. Tear testing is complex in its nature, and the value is not related directly to rubber performance as the results are dependent on consistency of sample preparation. Tear strength $T_S$ is relationship between the maximum force F (in N) and test piece thickness d (in mm). [18]

$$T_s = F/d \quad (1)$$

Hardness testing indicates if the rubber is resistant to indentation. Hardness is tested at logarithmic durometer scale Shore A, where higher values indicate harder material. It is simple, inexpensive, and fast but somewhat crude method with big variability and poor repeatability. It is recommended to be used as a quick and simple method to spot big differences in cured compound batches. [18] The hardness of vulcanized compound is measured to ensure that it does not vary too much, as hardness and stiffness alone can influence other properties [13].

Abrasion and wear resistance are important for the product's lifetime as most of the tires are replaced because they become worn out. Abrasion tests try to estimate the wear resistance of the tire in real conditions. Wear resistance is a complicated feature, and it is related to many other properties such as cut resistance, tear resistance, fatigue resistance, hardness, resiliency,

and thermal stability. The real lifetime of a tire is measured in years, but abrasion tests try to accelerate the process which leads to test not always corresponding to normal driving. [18]

### 2.2.3 Dynamic mechanical properties

Tires are used during both cold winters and hot summer days. The tire properties are not uniform in the whole operating temperature range. Dynamic mechanical properties reveal important data on compound behavior on different temperatures, amplitudes, and frequencies. There are two methods to study dynamic mechanical properties: dynamic mechanical analysis (DMA) that measures tensile modulus and rubber process analyzer (RPA) that measures shear modulus. The results from DMA and RPA seem to be similar and can be used analogously. [21] This thesis is focused on DMA.

Dynamic mechanical properties are the viscoelastic properties that are measured by applying sinusoidal (oscillatory) load and recording stress and strain of the sample. It used to evaluate dynamic properties like hysteresis which refers to loss of mechanical energy as heat. Storage normal modulus (E') is the component of normal stress in phase where higher value indicates stiffer material. Loss modulus (E'') estimates the energy lost as heat when the sample is deformed, where higher value means greater energy loss. Complex normal modulus (E*) expresses total resistance to deformation and it is calculated from E' and E''. Loss factor (tan δ) is the ratio of loss modulus to storage modulus, where higher values mean a more hysteretic compound. Phase angle δ expresses out-of-phase response to applied stress. [18]

$$|E^*| = \sqrt{(E')^2 + (E'')^2} \quad (2)$$

$$tan\delta = E''/E' \quad (3)$$

When evaluating tire performance, several values from DMA measurements can be used. In the industry, following values are used most commonly. Tan $\delta$ at 30°C and 60 °C are used for predicting rolling resistance where lower values are better. Tan $\delta$ at 0°C and -10 °C are used for predicting wet traction and ice traction respectively where higher values are better for both. $E^*$ at -20 °C is used to predict winter traction and E' at 30 °C dry handling where lower is better and higher is better respectively. [22]

### 2.2.4 Property indications

When comparing different compounds, it is important to know what the indication of each value is. Some properties have clear indications; for example abrasion resistance is pretty self-

explanatory, and others are backed by years of experiments in the industry, like lower tan $\delta(60°)$ indicates improved rolling resistance and treadwear when used in tread [23]. Other properties such as vulcanization characteristics are more related to processability and can be a trade-off between processing costs, safety, and quality.

Properties, their units, indications, and preferred values are summarized in the Table 2.

Table 2 Summary of tire compound properties

| **Property** | **Unit** | **Indication** | **Preferred value** |
|---|---|---|---|
| $M_L$ | dN.m | Processability | Balanced |
| $M_H$ | dN.m | Crosslink density | Balanced |
| $M_H$-$M_L$ | dN.m | Change in stiffness | Smaller |
| $t_{10}$ | min | Processing safety and flexibility | Longer |
| $t_{90}$ | min | Cure time, process safety | Balanced between shorter for manufacturing and longer for safety |
| Modulus 300 | MPa | Better reinforcement | Higher |
| Tensile strength | MPa | Resistance to breaking and deformation | Higher |
| Elongation at break | % | Related to tensile strength, ability to deform before breaking | Balanced between stiffness and strength |
| Hardness | ShA | Resistance to deformation, used for quality control | Constant between batches |
| Abrasion | g | Better wear resistance | Lower |
| Tear Strength | N/mm | Resistance to crack propagation | Higher |
| tan δ (-10 °) | - | Wet/ice grip | Higher |
| tan δ (0 °) | - | Wet/ice grip | Higher |
| E’ (30 °) | - | Dry handling | Higher |
| tan δ (60 °) | - | Rolling resistance | Lower |

# 3 Machine Learning

Traditional tire design relies heavily on experiment-based methods and the years of knowledge that has formed into industry. As testing new tire compounds and tires is time-consuming, machine learning could be utilized to speed up the process by predicting the properties of different tire recipes. Many tire manufacturers are using artificial intelligence and machine learning tools to improve their processes and simulations [4,24].

The relationships between different compound materials have proven to be complex and there is no universal equation for compounding to get the desired properties. It is extremely challenging to create a mathematical model between compounds and properties because there are countless variations in processing and raw materials [25]. Despite the complexity, machine learning has been successfully used for polymer materials design and discovery [26]. Regarding tire compounding, a machine learning model aims to find a relationship between the input materials and wanted properties which can help to optimize the recipe faster.

## 3.1 Machine Learning for Property Prediction

Despite their simple outlook, tires are very complex to produce with multiple changing parameters. Data-driven approaches have gained popularity in material research and tire development is no exception.

Artificial neural networks (ANN) have been used widely in predicting rubber properties. Schwarts [25] used ANN to predict rheometric properties of rubber compounds in early 2000's with average relative error of 11.3 %. Vijayabaskar et al.[27] studied predicting mechanical properties of different rubber formulations using ANN. They received average relative error of 11.8 % for modulus 300, and less than 5 % for the rest. They noted that number of training data was a key requirement for prediction quality of ANN. Wang et al. [28] studied predicting abrasion and mechanical properties of SBR based rubber composites using ANN with prediction accuracy of 96.0 %. Uruk et al. [29] studied using ANN and hybrid approaches of ANN to predict rheometric properties of rubber compounds. They got average mean percentage errors of less than 5 % for all the properties.

Other methods have been compared for property prediction too. Pang et al. [6] used multiple linear regression (MLR), ANN, support vector machine regression (SVR), and classification and regression trees to predict tan $\delta$ at 0 °C and 60 °C, tensile strength, and hardness of tire

tread composites and filler system using carbon nanotubes. They got smallest prediction errors of < 5 % for MLR. Huang et al. [5] used data from molecular dynamics simulations and predicted tensile strength of natural rubber blends using extreme gradient boosting (XGB) and compared the prediction results to MLR and SVR. They received $R^2$ values in the range of 0.94-0.98 for XGB, which was the best performing model. Machine learning has also been used for reverse engineering purposes where compound is predicted based on the properties. Román et al. [30] compared linear regression, response surface method (RSM), ANN, and Gaussian process regression (GPR) predicted compound formulation from target properties. They received prediction accuracy of 90 %, 97 %, and 100 % for RSM, ANN, and GPR respectively.

## 3.2 Regression

As the literature shows, many supervised machine learning methods are used to predict a variety of tire-related data. There are two types of supervised learning: regression and classification. In supervised learning, the data is divided into independent and dependent variables. Independent variables are the input values and dependent variables are the outputs that change based on the independent variables. Regression aims to predict a quantitative target value. In classification, the target variable is qualitative, and it assigns a class or a category to the observation. In this thesis, target properties have quantitative values. The goal is to create a model that can predict numeric output value based on the input variables. The used input values are proportions of compound materials and output variables are continuous values for properties.

Regression is widely used machine learning method for outputs with continuous values [31]. The problem at hand is multi-output regression where the output is 2D array. Zhang and Ling [32] have demonstrated that it is possible to reach high precision in regression tasks even when using small dataset in material science.

Most supervised models are built on assumption that data points are drawn independently from the same distribution. This kind of data is called independent and identically distributed (IID) [33]. This assumes that the future (unseen) data is drawn from the same distribution as the data that was used for the training. If the assumption of IID is violated, the model does not function correctly which lead to degradation in performance. When the unseen data has a change in the distribution of a single feature or combination of features, it is called dataset

shift. This dataset shift often happens in real-world datasets. [34] This has been studied a lot on classification, but it holds true for regression tasks also.

IID assumption explains why extrapolation is difficult for most of the models. Models are tuned to excel in interpolation i.e. predicting data points within the existing dataset. Data points outside the range of the dataset may not belong to the same distribution and cause a problem for a model. This needs to be addressed when the model is trained, and the performance of the model is evaluated.

## 3.3 Training and Model Tuning

Machine learning model needs to be trained with a dataset before it can be utilized. In supervised learning, a dataset is given to the model, and the model tries to learn the underlying relationships in the data to predict a target value. Usually, a part of the dataset is left as a test data to see how well the model can generalize its learning to new cases.

Many ML models have tunable parameters. Changing these parameters changes the behavior of the model. Default parameters usually give a good start but in most cases the parameter finetuning benefits the model quality drastically. Models with different parameters are compared to each other using evaluation metrics. Evaluation metrics are explained in the next paragraph. Model specific parameters are presented later in the chapter.

## 3.4 Evaluation

Machine learning models need effective evaluation methods to assess how well they are doing the given task. There is no one model that dominates on all datasets. The precision of ML models depends on several factors such as dataset size, data quality (outliers, noise, errors in data), and model parameters. Knowing the type and quantity of the error helps to select the most suitable model.

Cross-validation is commonly used to estimate how well the model will generalize on unseen data on average. For the evaluation, the dataset is divided into test and training sets. In cross-validation, the data is resampled multiple times into different test and training sets to handle the variability in the data. [35] Training set is used to train the model. Samples in the test set are then given to model, and the model predicts their value. How close these values are to the original values determines the precision of the model.

Many different metrics are used to evaluate the regression model [31]. All the metrics have their own strengths and weaknesses on different datasets and therefore multiple metrics are commonly used to evaluate the model. Commonly used regression metrics are coefficient of determination ($R^2$), mean absolute error (MAE), and mean root squared error (RMSE).

According to Chicco et al. [31] MAE, RMSE, and $R^2$ can be calculated with following formulas:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (5)$$

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(\bar{y} - y_i)^2} \quad (6)$$

where $\hat{y}_i$ is predicted i$^{th}$ value, $y_i$ is the actual i$^{th}$ value, $\bar{y}$ is mean of the observed data, and N is the number of datapoints.

MAE gives a generic performance metric for the performance whereas RMSE gives more weight to outliers. The best value for both MAE and RMSE is 0 and the worst value is $+\infty$. This means it is more difficult to interpret the MAE or RMSE value alone if the value is for example 0.8 as it depends on the distribution of the target value. The best value for $R^2$ is 1, the worst value $-\infty$, and the value 0 indicates that the regression model explains none of the variability. [31] $R^2$ score of 0 can be received by guessing the mean for all the values, and negative values are worse than guessing the sample mean.

As RMSE and MAE are sensitive to the scale and unit of values it makes comparison of different target properties difficult. Thus, normalized versions of them are used in this thesis for better comparison between the target variables. $R^2$ is dimensionless metric and has been used to compare different properties [5,30].

There is no agreement in using dimensionless metrics in literature. Many studies have used relative error as a dimensionless metric in evaluation as discussed in chapter 3.1. Tofallis [36] has pointed out few problems in using relative error and mean absolute percentage error (MAPE). The problems come from that they divide absolute error with the true value.

$$relative\ error = \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (7)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100\ \% \quad (8)$$

If the actual value is zero or close to zero this causes error values to approach infinity. In addition, the formulas are asymmetric where interchanging $\hat{y}_i$ and $y_i$ leads to different errors.

If the dataset has actual values that are zero or close to zero, another approach is taken and MAE and RMSE are normalized for the comparison. Based on the dataset characteristics, the mean, the range, the standard deviation, or the interquartile range can be used in normalization. The range is difference between maximum and minimum values, and it is used here as it gives clear interpretation between differently distributed target values. The value is given as a percentage to show how many percents the error is in the data range. The normalized versions of MAE and RMSE are the following:

$$NMAE = \frac{MAE}{y_{max} - y_{min}} * 100\ \% \tag{9}$$

$$NRMSE = \frac{NRMSE}{y_{max} - y_{min}} * 100\ \% \tag{10}$$

where $y_{max}$ is the highest value of the variable and $y_{min}$ is the lowest value of the variable.

Kendall rank correlation coefficient (also known as Kendall's $\tau$) is a statistical method to compare similarity of ranks between two set of ranks. It checks the concordance of all the possible pairs in the data. Values close to 1 indicate strong agreement and values close to -1 strong disagreement in ranks. Values close to 0 indicate there is no correlation between the rankings. Kendall's $\tau$ can be calculated between two lists x and y using the following formula [37]:

$$\tau = \frac{P - Q}{\sqrt{(P+Q+T) * (P+Q+U)}} \tag{11}$$

where P is number of concordant pairs, Q is number of discordant pairs, T is number of tied pairs only in x, and U the number of tied pairs only in y.

Kendall's $\tau$ is usually reported with the p-value that tells the statistical significance of the observation. If the p-value is small ($p < 0.05$) the observed $\tau$ is statistically significant. With large p-values ($p > 0.05$) the observed correlation can be due to random chance.

Kendall's $\tau$ is not traditionally used as an evaluation metric in material science machine learning. Here Kendall's $\tau$ is used because it is interesting to see if the models can keep the ranking of the original data. In tire development, it is informative to know if the new compound is performing better or worse than reference compounds and the absolute value is less important.

## 3.5 Models

Because using ANN is computationally expensive and usually requires a large training dataset, several other methods are selected for this study. The selected models work well on smaller datasets. The basic principles of those models are explained below.

### 3.5.1 Linear Regression

Linear Regression (LR) is a simple statistical technique that is widely used to model relationship between the variables. Due to its simplicity, linear regression is easy to implement and interpret.

A multiple linear regression model is an extension of LR that forms a relationship between a dependent variable and multiple independent variables. The formula can be written as follow, where i is the $i^{th}$ observation, and k is the number of the features:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (12)$$

where $y_i$ is the dependent variable, $x_{ij}, j = 0,1,2 \ldots k$ are independent variables, $\beta_o$ is the intercept, $\beta_j, j = 0,1,2 \ldots, k$ are the regression coefficients, and $\epsilon_i$ is the error of the $i^{th}$ observation. [38]

To get the best possible predictions $\hat{y}$, least squares method is used to try to find coefficients estimates $\hat{\beta}_j, j = 0,1,2 \ldots, k$ for LR that minimize the sum of squared residuals (RSS) [35]:

$$RSS = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_k x_{ik}\right)^2 \quad (13)$$

where N is the number of training samples.

Figure 1 shows how the model tries to find the best fit for the data. The green line represents better fit, and green arrows corresponding residuals. Respectively, orange line represents a worse fit and orange arrows corresponding residuals. Longer orange arrows represent larger residuals.

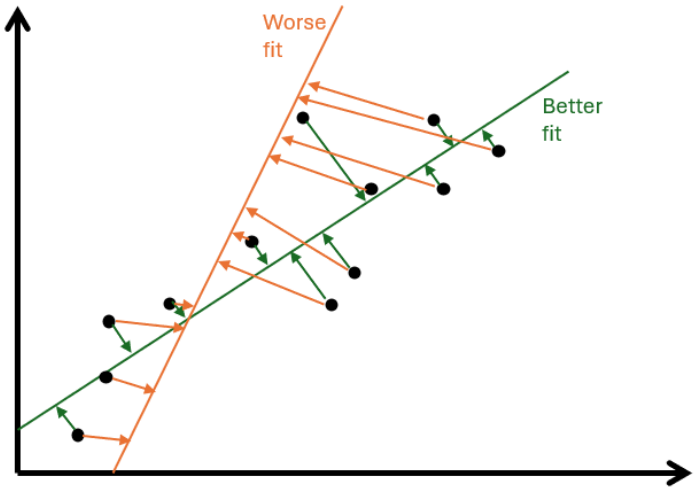


Figure 1 Linear Regression with two fits for the same data points. Green line has smaller errors and is consider better.

There are several disadvantages related to linear regression. As the name suggests, the relationship between dependent and independent variables should be linear. Ideally, residuals are independent and normally distributed, and there is no collinearity between predictor variables. Linear regression is also sensitive to outliers. [35] All these factors cause problems with the precision of linear regression. Due to complex nature and underlying principles of rubber compounding, it is likely that linear regression is not the optimal model for compound property prediction. LR is used here as a baseline algorithm as it has many advantages like simple implementation, good explainability, and computational efficiency.

### 3.5.2 Ridge Regression

The performance of linear regression weakens when input features are correlated with each other. Hoerl and Kennard [39] proposed ridge regression (RR) for nonorthogonal problems. RR is similar to linear regression, but it utilizes the regularization term to shrink regression coefficients towards zero. The shrinking is used to improve the fit by adding a little bit of bias to reduce the variance. Instead of just RSS, ridge regression tries to minimize RSS and a shrinkage penalty:

$$\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_k x_{ik}\right)^2 + \lambda\sum_{j=1}^{k}\beta_j^2 = RSS + \lambda\sum_{j=1}^{k}\beta_j^2, \qquad (14)$$

where $\lambda \geq 0$ is tuning parameter that is tuned to suitable value for the dataset and k is the number of features. The penalty affects all the other coefficients except intercept $\beta_0$. [35]

Figure 2 shows how shrinkage parameter works by introducing some bias to the model.

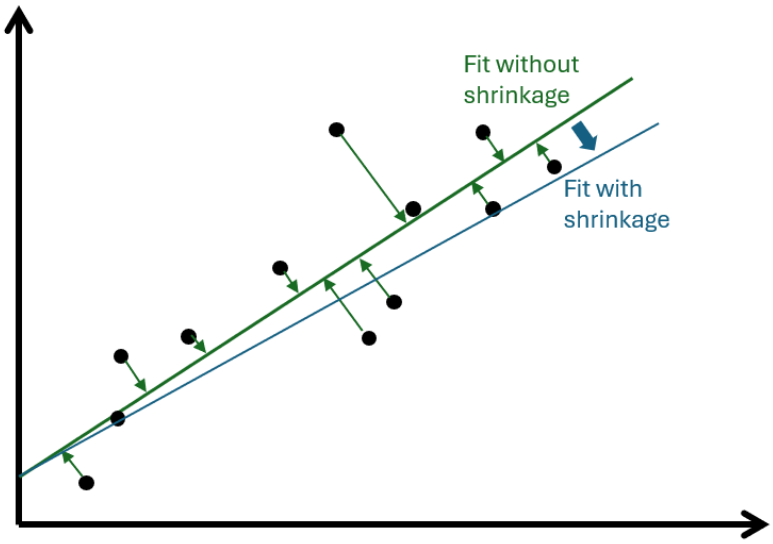


Figure 2 Effect of shrinkage parameter. The coefficient is moved closer to zero.

When there are multicollinearity in the features, RR gets lower values for mean squared error due to the $\lambda$ and can thus be more precise than LR. As $\lambda$ introduces some bias to the model, RR performs well when least squares method results in high variance in the model that can be reduced with a small increase in bias.

### 3.5.3 Support Vector Regression

Support vector regression (SVR) is a variant of support vector machine that is commonly used for classification. SVR tries to fit a hyperplane into the training data so that the data deviates no more than $\varepsilon$ of the plane. SVR can use linear and non-linear kernels to model data.

The following description of the principles of SVR is based on references [40–42]. SVR is a linear method, but it can perform nonlinear regression with $\Phi(X_i)$ mapping. SVR seeks a function

$$f(X_i) = \langle w, \Phi(X_i) \rangle + b \tag{15}$$

where $\langle \cdot, \cdot \rangle$ is the dot product between the coefficients w and transformed features and b is a constant bias term. Function is based on data $(X_1, y_1), \ldots, (X_i, y_i) \in R^k \times R$ with k features. w and b are optimized to minimize an objective function that combines the norm of the weights $\|w\|^2$ and empirical risk function based on Vapnik's $\varepsilon$-sensitive loss function:

$$L_\varepsilon[y_i - f(X_i)] = \begin{cases} o, \ if \ |y_i - f(X_i)| \leq \varepsilon \\ |y_i - f(X_i)| - \varepsilon, \ otherwise \end{cases} \quad (16)$$

Thus, only the values greater than $\varepsilon$ contribute to the loss function.

Next in SVR, the parameters are optimized to minimize objective function using slack variables $\xi_i, \xi_i^*$:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (17)$$

$$\text{subject to} \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

where $C > 0$ is a constant that determines the trade-off between flatness of $f(x)$ and how large deviations from $\varepsilon$ are tolerated.

Minimizing can be solved by using Lagrange multipliers $\alpha_i, \alpha_i^*$

$$\begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j \rangle \\ -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \quad (18)$$

$$\text{subject to } \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C]$$

Linear kernel is the simplest kernel, and it is used when data is linearly separable.

$$K(X_i, X_j) = \langle X_i, X_j \rangle \quad (19)$$

A commonly used kernel for non-linear problems is the Gaussian radial basis function (RBF):

$$K(X_i, X_j) = \exp\left(-\gamma \|X_i - X_j\|^2\right) \quad (20)$$

where $\gamma$ is the constant parameter of the kernel.

The regression function can be rewritten with the kernel function, Lagrange multipliers, and the bias term:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K\left(X_i, X_j\right) + b \quad (21)$$

Lagrange multipliers and bias are trained from the data.

Figure 3 shows how SVR can fit a hyperplane to non-linear data. The black solid line is the hyperplane; blue dashed lines represent support vectors that form the $\varepsilon$ tube around the hyperplane.

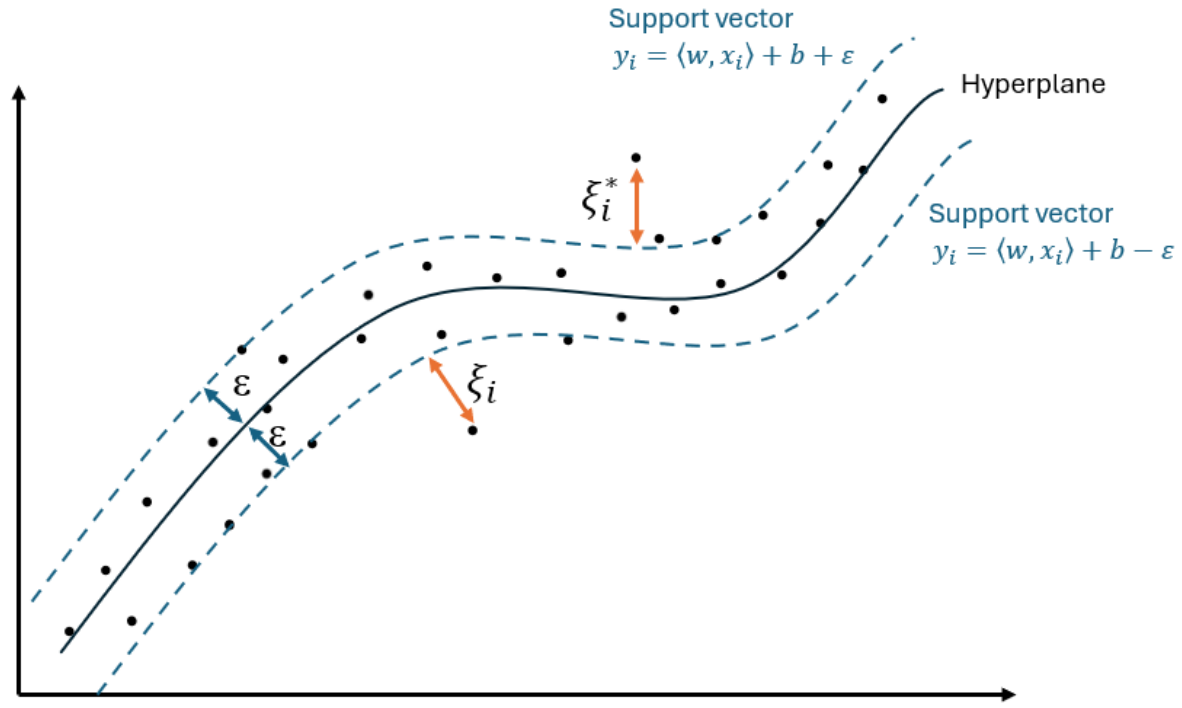


Figure 3 Support vector regression working principle and terms visualized.

Using SVM includes hyperparameter tuning which is crucial for the model. $\varepsilon$specifies the deviation from the plane. C is regularization parameter to control bias-variance trade-off. As SVR is a kernel-based method, the user must select a kernel and kernel-specific parameters such as $\gamma$ in case of RBF.

There are several advantages to using SVR that come from its working principles. It is a kernel-based method that can be used for high-dimensional data, and it can handle non-linear data. The model can be controlled with hyperparameters C and $\varepsilon$. It can also handle outliers well. Disadvantages include sensitivity to hyperparameter tuning and less interpretability. SVR also requires standardization of the input data as it is sensitive to distance, and features with larger range would dominate in the calculations.

### 3.5.4 Kernel Ridge Regression

Kernel Ridge Regression (KRR) is a special case of SVR, but it uses different loss functions. While SVR uses $\varepsilon$ tube, KRR uses loss function presented in above in ridge regression. KRR tries to find a function $f$ that minimizes the empirical risk [43]:

$$\min_{f \in \mathcal{F}} \left( \sum_{i=1}^{N} \left( y_i - f(x) \right)^2 + \lambda \|f\|_{\mathcal{F}}^2 \right) \quad (22)$$

where $\mathcal{F}$ is the Reproducing Kernel Hilbert Space (RKHS) with selected kernel. The benefit of using RKHS is that it allows linear models like ridge regression model non-linear relationships in the data by using higher dimension.

In addition to Gaussian kernel (RBF), its variant Laplacian kernel can be used. It uses Manhattan distance to calculate the distance of the input vectors. [44] The use of absolute values makes it less sensitive to outliers compared to RBF.

$$K(X_i, X_j) = \exp\left(-\gamma \left\| X_i - X_j \right\|_1\right) \tag{23}$$

Figure 4 illustrates how moving the data into higher dimensions can help to fit a linear model to the data.

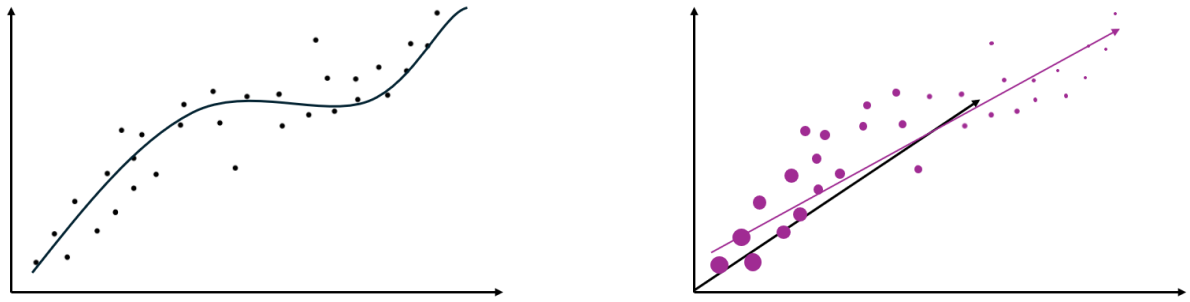

Figure 4 The data is moved to higher dimension to help to find a line that can fit non-linear data

### 3.5.5 Random Forest Regression

Random Forest (RF) is an ensemble model of decision trees. Random forest consists of B individual tree predictors $T_b$. Trees are grown using subsets of the original dataset and the output result is the average of B trees:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x; \Theta_b) \tag{24}$$

where B is the number of trees and $\Theta_b$. is independent random vector for $b^{th}$ value. The randomization aims to low correlation between the trees which makes RF more robust. [45]

Figure 5 gives an example of a regression tree with 4 features. The tree is predicting a target value based on the features. In this case, the tree ended up to value 1.3. In case of random forest, multiple trees are created, each tree makes its individual decision, and the final value is average of all the trees.

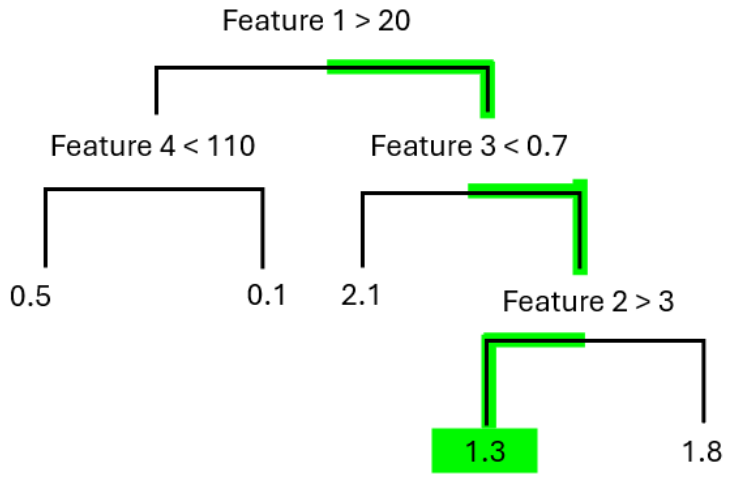


Figure 5 Single regression tree

In random forest, important hyperparameters are the number of trees in the forest and the number of features in the subset. Number of leaf nodes (number of answers the tree can give) and maximum depth (maximum number of features to be used to make a prediction) can be optimized to get best performance. [46] In the figure 5, the number of features is 4, number of lead nodes is 5, and maximum depth is 3.

RF can handle non-linear and high dimensional data, and it can provide information on feature importance. The selection of random trees makes it less prone to overfitting and outliers. [45] A single tree can be easily interpreted but multiple trees become quickly more difficult to interpret.

### 3.5.6 Gradient Boosting Regression

Gradient Boosting Regression (GBR) is another ensemble method that uses decision trees for predictions. In GBR, trees are built sequentially. Each new tree adds information to the previous trees. The algorithm behind GBR starts with $\hat{f}(x) = 0$ and residuals $r_i = y_i$ for all i in the training set. Then a tree $\hat{f}^b$ is fitted to the training data and the model is updated by $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$ where $\lambda$ is the shrinkage parameter. The process is repeated for $b = 1,2, \dots B$ times. This results in boosted model [35]:

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x) \qquad (25)$$

Figure 6 has three example trees for GBR. The first tree is the starting tree, and the following trees change the value by factor of $\lambda$ to get closer to the real value.

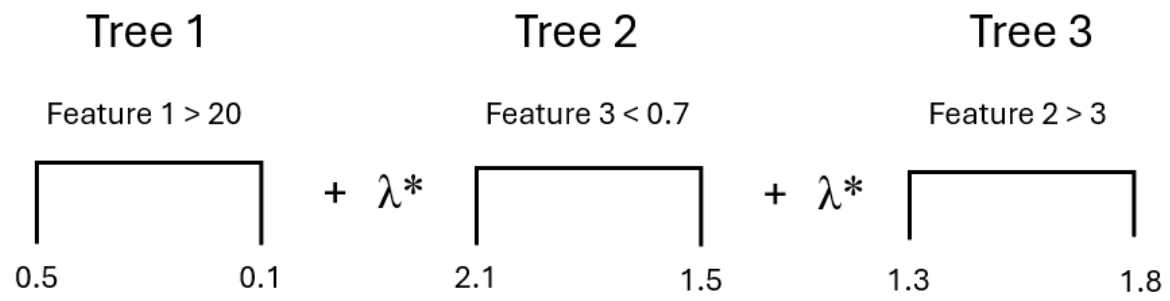


Figure 6 GBR with three example trees.

Like RF, GBR is also tuned for the number of trees. The shrinkage parameter $\lambda$ plays an important role in how much each tree affects the model. If $\lambda$ is very small typically a large number of trees is needed. The depth of the trees controls the complexity of the ensemble. The decisions made by the tree are called splits. Typically shallow trees are preferred where the number of splits is one [35], meaning only one decision per tree, as shown in the figure 6.

As GBR builds gradually by minimizing the residuals, it can improve on its weak areas, and it works well on heterogeneous data. However, tuning GBR requires effort and the prediction ability for new cases can be very poor if there are new features (materials) as the model is optimized for existing features.

### 3.6 Overfitting and underfitting

Overfitting and underfitting are common problems in machine learning. Usually, model is aimed to be robust with good generalization ability. In overfitting, the model learns the training data too well and produces a complex model that has poor generalization for unseen data. In underfitting, the model is too simple and cannot learn the relationships from the data. Figure 7 shows three fits. The balanced fit is a compromise between overfitted and underfitted versions and is likely the most robust when new data is fitted to the model.

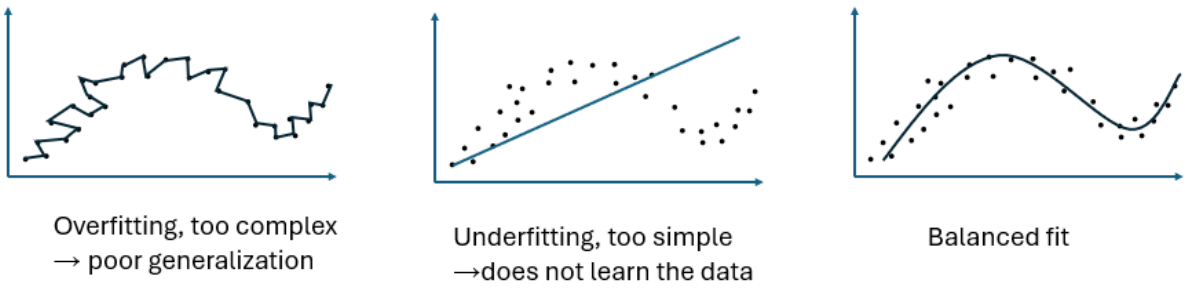


Figure 7 Overfitting, underfitting, balanced fit of same data points

Overfitting and underfitting are often result of bias and variation in the data. Effect of bias and variance are demonstrated in Figure 8. In ideal case, the data has low bias and low variance but that is not the case in many real-world datasets.

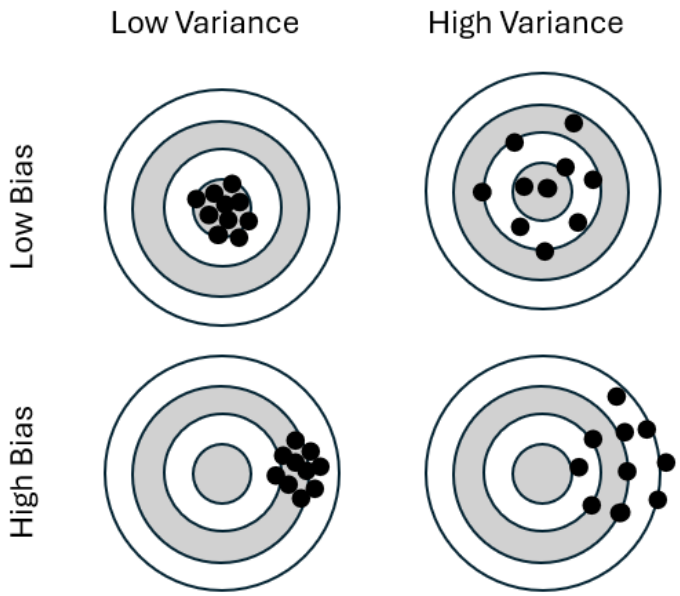


Figure 8 Bias and Variance

Poorly selected features or sampling errors can result in bad model. Small dataset can lead to underfitting and a large bias of prediction whereas high degree of freedom can lead to high variance and overfitting [32].

# 4 Methods

The data was analyzed, modeled, and visualized using Python programming language and following open-source libraries: Python (version 3.13.2), joblib (version 1.4.2), matplotlib (version 3.10.1), numpy (version 2.2.4), optuna (version 4.4.0), pandas (version 2.2.3), scikit-learn (version 1.6.1), scipy (version 1.15.2), and seaborn (version 0.13.2).

## 4.1 Data

The dataset set used in this thesis is small. Data was combined from nine laboratory experiments with a total of 86 samples. The nine experiments were used as groups in this thesis. Experiments were done using ASTM and ISO standard test methods to make it easier to compare the data between the experiments.

Material PHR values were used as input variables and if the material was not used it was marked as zero. There were 81 different materials used as features. Each compound consisted of 3 to 20 different materials with median of 15. Different materials were rubbers, fillers, sulfur, vulcanization activators and accelerators, antidegradants, and process aids. The number of materials in this study is larger, and the samples are more varied with less samples than most previous studies presented earlier. This reflects the nature of real-life experiments where data is scarcer and more diverse.

Eleven properties were chosen as target variables. One experiment was missing dynamic mechanical properties, so 66 samples were used for testing those properties. Target variables were values for vulcanization characteristics: maximum torque ($M_H$), minimum torque ($M_L$), difference of maximum and minimum torque ($M_H$-$M_L$), and cure time ($t_{90}$); mechanical properties: tensile strength, modulus 300, elongation at break, and hardness; dynamic mechanical properties: tan$\delta$ (0°C), tan$\delta$ (60°C), and E' (30°C). Same properties have been studied in previous studies as they are commonly used in rubber research and development.

Table 3 shows how the data was collected and arranged into one table for modeling. The values in the table are fictious and for the visualization purpose.

Table 3 Data structure representation using fictious values.

| Type | Material | Material | … | Material | Group | Property | Property | … | Property |
|---|---|---|---|---|---|---|---|---|---|
| Column | 1 | 2 | … | 81 | | $M_L$ | $M_H$ | … | E' (30°C). |
| Sample 1 | 0 | 10 | … | 2.5 | 1 | 2.88 | 14.47 | … | 15.1 |
| Sample 2 | 0 | 10 | … | 3 | 1 | 2.69 | 13.52 | … | 11.45 |
| Sample 3 | 3 | 12 | … | 0 | 1 | 1.35 | 16.34 | … | 13.05 |
| … | … | … | … | … | … | … | … | … | … |
| Sample 86 | 0 | 0 | … | 1.4 | 9 | 3.29 | 21.78 | … | - |

## 4.2 Data preprocessing

First all the experiments were collected in a single table and PHR values were used as input data with zeros when material is not part of the recipe. One column had experiments numbers which were used for visualization and interpretation purposes and in LOGO cross-validation, but they were not part of the training data.

Only few of the materials were left out. Those materials were present in only one experiment and fully correlated with some other material in the experiment, and thus they provided no additional value. Other highly correlating materials were kept in data. Marking unused materials as zero increased correlation for many of the materials and the initial data analysis showed that other highly correlating materials usually included some valuable information about the experiment that would be lost if one of the highly correlating materials were removed.

Figure 9 shows the correlation matrix for all the features. Each box represents the strength of correlation between two features where darker color means stronger positive or negative correlation.

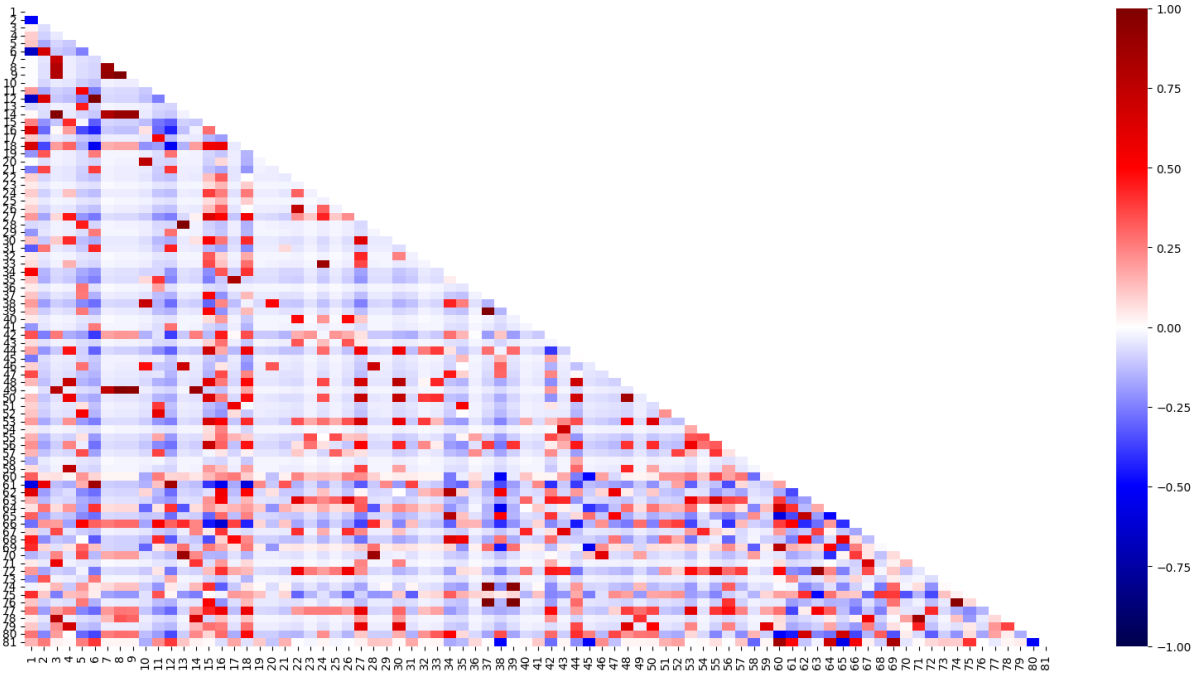


Figure 9 Correlation matrix of the selected features. Numbers represent the column number. Each square represents the strength of correlation between each feature pair. Dark red indicates strong positive correlation and dark blue strong negative correlation. Light values indicate low correlation.

Input features were scaled using StandardScaler function from scikit-learn for all methods. Tree-based methods are not sensitive to scaling, but using scaling for them too makes sure that all the models had the same input values. StandardScaler standardizes features by centering them around zero and to have standard deviation of 1 and each column is centered and scaled independently [47]. Standard score z of training sample x is calculated as follows:

$$z = (x - u)/s \tag{26}$$

where u is the mean of the training samples and s is the standard deviation of the training samples.

To avoid data leakage during cross-validation, StandardScaler was first fit only to the training data. Then both training and test data were transformed using same scaler. This is to avoid data leaking during the training which can give overly optimistic results and to evaluate how the models perform on unseen data.

StandardScaler is sensitive to outliers. In this dataset the data is naturally sparse which makes it look like some features could have outliers as most of the samples include only a limited amount of materials. Scaling the features can improve the prediction precision as effect of a tire material are often disproportionate to their amount in compound. Especially, SVR and KRR benefit from scaling as all kernel-based methods are based on distance and without

scaling, the feature with greater numeric range would dominate in the computation [48]. Results were not scaled but normalized versions of RMSE and MAE values were used for easier comparison.

## 4.3 Data modeling

Data was modeled using Python and open-source library scikit-learn. Selected models are linear regression, ridge regression, kernel ridge regression, support vector regression, random forest regression, and gradient boosting regression. A separate model was trained for each property to ensure the optimal performance. The size of the dataset affected directly the chosen methods. Even though ANNs are widely used in previous studies, they were left out as they require large datasets.

As the number of samples is small, the data was divided into testing and training set using leave-one-out cross-validation (LOOCV) and leave-one-group-out cross validation (LOGOCV). In LOOCV, each value is used as a test value against others that act as training set which is useful for testing the model performance for new unseen data points. The test error is the average of all iterations. LOGOCV is a special case of LOOCV where one data group is tested against others. LOGOCV is suitable for testing the model's ability to predict properties for new unseen groups or individual data points from unseen groups. Experiments were used as groups in LOGOCV.

Due to simplicity of linear regression, preprocessed data was fitted to the regression without extra steps.

The following steps were performed to all the models with hyperparameters. Grid search cross-validation was used for the scaled data to find the best parameters. RMSE was used to select the best hyperparameters. Hyperparameters were separately tuned for each property and LOOCV and LOGOCV. Selected hyperparameters were then used to train the training data and get all the metrics using both LOOCV and LOGOCV.

Ridge regression was used to counter the nonorthogonality in the dataset. Parameter $\lambda$ was tuned to make regression more resistant to orthogonality. Kernel ridge regression allows using nonlinear space. For KRR, two kernels, RBF and Laplacian, were evaluated alongside $\lambda$ and $\gamma$. Using linear kernel would give the same results as using regular Ridge regression. SVR

was another kernel-based method. Linear and RBF kernels were chosen for evaluation with hyperparameters C, $\varepsilon$, and $\gamma$.

Random forest and gradient boosting both have parameter random_state in scikit-learn package. This was set to be constant to make evaluation of different iterations easier as the random split can affect the end results. RF was tuned for number of trees, number of features, tree maximum depth, minimum samples per leaf, and bootstrap. GBR was tuned for number of trees, learning rate, minimum samples per leaf, maximum number of features, and tree maximum depth.

## 4.4 Evaluation

After the hyperparameters were selected models' test errors were evaluated with $R^2$, NRMSE, and NMAE. Kendall's $\tau$ was used to evaluate how well the models can keep the original ranking of the data. In addition, actual and predicted values were plotted to evaluate the distribution and location of the errors. This helped spot how much variance and bias there is in the model.

# 5 Results

This chapter discusses the results of machine learning modeling. The prediction metrics were calculated for six different models using two different cross-validations LOOCV and LOGOCV. Used metrics are $R^2$, NRMSE, NMAE, and Kendall's $\tau$. Models trained using LOOCV performed significantly better than LOGOCV. The models derived from LOGOCV demonstrated poor predictive performance that is not sufficient for practical applications. Thus, this chapter will mostly focus on LOO results unless otherwise specified.

Separate models were trained for each property to get optimal performance because the properties have different background principles and their scales and units varied. First, the hyperparameters were tuned and the best combination of hyperparameters was selected based on the RMSE value. Then, all the metrics were evaluated using the model with best hyperparameter combination. All the models benefitted from hyperparameter tuning compared to default parameters in scikit package. Overall, there was no single best model that would surpass other models in all properties and the results varied for each property.

## 5.1 Target properties

In total, there was 11 target properties: maximum torque ($M_H$), minimum torque ($M_L$), difference of maximum and minimum torque ($M_H$-$M_L$), cure time ($t_{90}$), tensile strength, modulus 300, elongation at break %, hardness, tan$\delta$ (0°C), tan$\delta$ (60°C), and E' (30°C). The distributions of the properties varied. Distribution can affect how easy the prediction is and how complex model is needed. Symmetrical distribution with low standard deviation is generally easier than asymmetric with large standard deviation.

Figure 10 shows the distribution and coefficient of variation (CV) for each property. CV measures relative variability in the data. Using CV allows comparison of variation for targets with different scales as their means and standard deviations are not directly comparable. CV is calculated by dividing the standard deviation by mean:

$$CV = \frac{s}{\bar{y}} * 100\ \% \tag{27}$$

where s is the standard deviation of the samples and $\bar{y}$ is the mean of the samples. Lower CV values indicate that the sample values are located closer to the mean and thus easier to predict whereas higher CV values indicate more difficulties in the prediction.

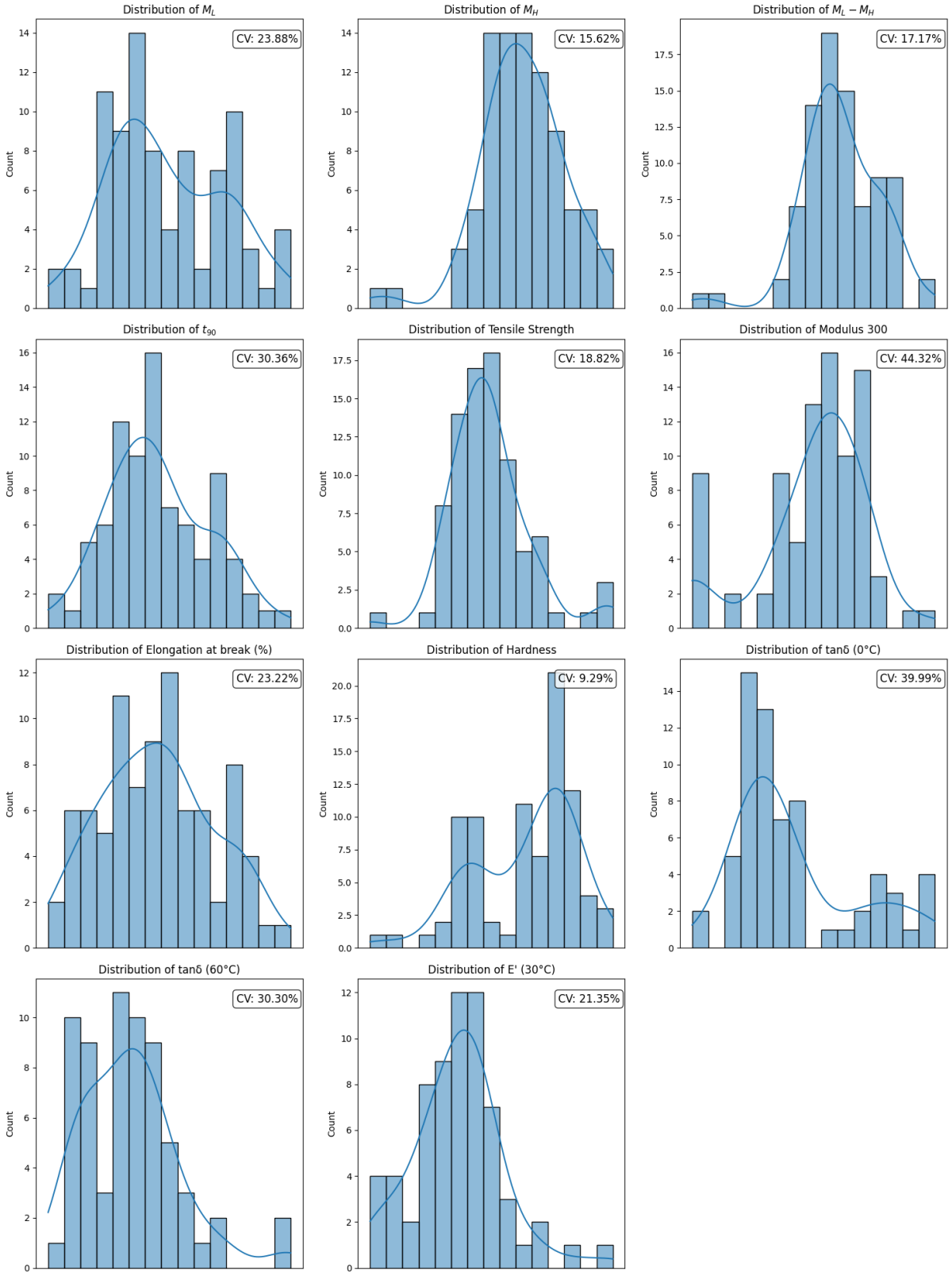


Figure 10 Distributions of the target values. Coefficient of variation values are shown in the upper right corner of each subplot.

Properties show diversity of distributions and CV values. As each of the properties has unique distribution, it is reasonable to treat them individually. Hardness has the lowest CV of 9.29 %

and modulus 300 has the highest of 44.32 %. Based on the CV, hardness should be the easiest to predict. None of the models seem to follow normal distribution but are more or less skewed which can increase the error for the values on the tail side of the distribution. Some distributions, such as hardness and tan $\delta$ (0°C), show signs of bimodality which can be more difficult to model.

There are few very high or low values in most of the distributions that are separated from the rest. These kinds of extreme values are usually more difficult to predict accurately, and they easily increase RMSE values. This could indicate outliers that differ from the rest of the values. They can also be special cases that need to be studied more carefully to find the reason for the deviation, but it is not within the scope of this thesis.

The varying distributions can cause problems for the predictions. They can indicate bias in the data. The size of the dataset is small which can obstruct the formation of distribution leaving some areas underrepresented and thus more difficult to predict.

## 5.2 LOOCV

The metrics agreed on the order of the models in LOO. On all metrics, KRR performed best and linear regression worst among the models on average. GBR was second, SVR third, RF fourth, and RR fifth. Only Kendall's $\tau$ showed some difference with KRR being first, GBR second, RR third, SVR fourth, RF fifth, and LR last. p-values for Kendall's $\tau$ were all under 0.01 except for two models in linear regression. This indicates that the $\tau$ values are statistically significant. Highest $\tau$ was 0.774 using SVR for $M_L$ so none of the models were particularly good at keeping the original sequence of the samples but mostly better than randomly ranking the values. In general, the residuals for LOO models are evenly distributed. This can be observed from the actual vs. predicted plots presented later in the chapter.

Table 4 shows the $R^2$ scores for models trained using LOOCV. Tables for NMAE, NRMSE, and Kendall's $\tau$ can be found in appendix. The highest score is shown in bold in each row. Cells with score higher than 0.8 are highlighted with green. The value of 0.8 was selected arbitrary to point out the best performing models. As shown on the table, one model does not fit all properties but difference in $R^2$ score are relatively small except for LR.

Table 4 R2 values for the models cross-validated with LOO. The highest value in each row is shown in bold. Cell with values over 0.8 have green background.

| | LOO $R^2$ | | | | | |
|---|---|---|---|---|---|---|
| **Target** | **LR** | **RR** | **KRR** | **SVR** | **RF** | **GBR** |
| **$M_L$** | -24.3 | 0.773 | 0.796 | 0.807 | 0.819 | **0.824** |
| **$M_H$** | -184.2 | 0.724 | **0.771** | 0.736 | 0.738 | 0.750 |
| **$M_H$-$M_L$** | -202.0 | 0.692 | **0.755** | 0.707 | 0.739 | 0.752 |
| **$t_{90}$** | -138.8 | 0.811 | **0.824** | 0.820 | 0.804 | 0.813 |
| **Tensile Strength** | -158.4 | 0.797 | **0.816** | 0.774 | 0.779 | 0.784 |
| **Mod 300** | -72.4 | 0.173 | 0.350 | **0.352** | 0.236 | 0.282 |
| **EB %** | -214.5 | 0.636 | 0.689 | 0.642 | 0.732 | **0.738** |
| **Hardness** | -175.7 | 0.875 | 0.876 | **0.883** | 0.813 | 0.810 |
| **tan δ (0°C)** | -4.3 | 0.796 | 0.823 | 0.812 | **0.852** | 0.833 |
| **tan δ (60°C)** | -48.4 | 0.777 | **0.852** | 0.799 | 0.772 | 0.848 |
| **E' (30°C)** | -127.5 | 0.619 | **0.624** | 0.573 | 0.423 | 0.519 |
| **Average** | -122.8 | 0.698 | **0.743** | 0.718 | 0.701 | 0.723 |
| **Rank** | 6 | 5 | **1** | 3 | 4 | 2 |

There are multiple possible explanations why LR is not suitable for the data. It simply might be too simple algorithm to model the data. LR is not usually good on high dimensional data where feature correlate with each other. On the other hand, RR, which is very close to LR mathematically, received significantly better results than LR. In RR, regularization term can reduce overfitting, reduce noise, and handle multicollinearity in the data. As the samples come from different experiments, some noise is inevitable. Also, the sparse nature of the data makes it seem like there are correlation between the data points even when closer analysis proves otherwise.

Nevertheless, RR came short compared to kernel and tree models with few exceptions, so it is likely that there are non-linear relationships in the data that linear models cannot express. Unlike RR, kernel and tree models can model nonlinear dependencies in the data. Kernel methods can turn linear algorithm into non-linear one. Tree methods can capture non-linear dependencies without mathematically modeling them. GBR has typically performed better than RF in the literature which was the case here also.

## 5.3 LOGOCV

Models created using LOGOCV were ineffective in predicting properties. The models only showed very little predictability. In LOGO models, the groups are more concentrated on certain prediction value and largely over- or underestimated. The best model was RF predicting $t_{90}$. Figure 11 shows that even if some data points are predicted correctly, most of the predictions are far from the perfect fit and the data is not formed around the perfect fit but rather vertically. The model uses much smaller range to make predictions than the actual range and overestimates some experiments and underestimates other experiments. The problems with the experiment groups is well visualized in Figure 11. The model predicts similar values for the samples from the same experiment group meaning it cannot distinguish the samples in new experiment.

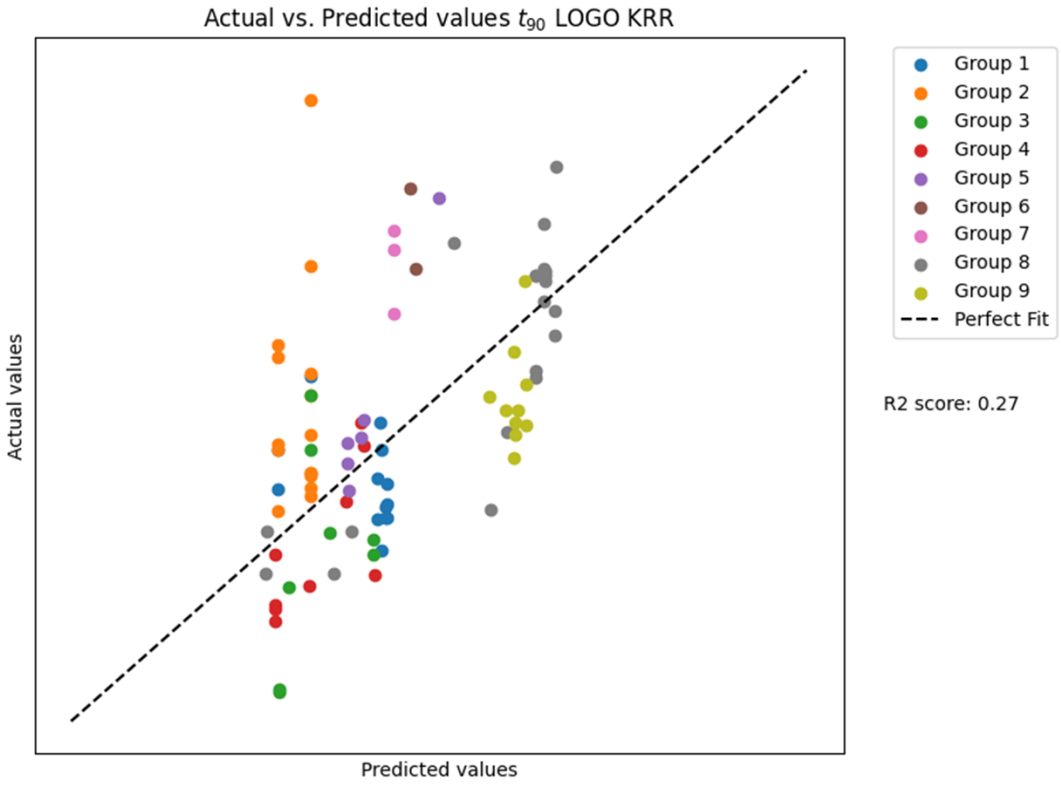


Figure 11 Actual vs. predicted values for $t_{90}$ using random forest model and LOGOCV. Many data points are far from the perfect fit.

Similarly to LOO, most of the metrics aligned on LOGO but Kendall's $\tau$ gave different ranking order. The p-values for $\tau$ were quite high in many cases which indicates that the calculated $\tau$ values are not statistically significant, and this metric is likely not reliable for LOGOCV. On other metrics, RF had the best rank on all metrics followed by GBR and SVR, and linear regression had lowest ranking. RF has inbuilt randomization in its algorithm. This

can be beneficial when predicting completely unseen samples with new materials as the model does not fit the training data too carefully. $R^2$ values were very low in general which means that the models could not explain the variance in data. Values close to 0 indicate that the performance is no better than random guessing and negative values indicate even worse performance.

Table 4 shows the $R^2$ scores for models trained using LOGOCV. Tables for NMAE, NRMSE, and Kendall's $\tau$ can be found in appendix. The highest score is shown in bold in each row.

Table 5 $R^2$ values for the models cross-validated with LOGO. The highest value in each row is shown in bold.

| | LOGO $R^2$ | | | | | |
|---|---|---|---|---|---|---|
| **Target** | **LR** | **RR** | **KRR** | **SVR** | **RF** | **GBR** |
| **$M_L$** | -256 | -0.425 | -0.424 | **0.045** | -0.297 | -0.227 |
| **$M_H$** | -234 | -0.311 | -0.487 | -0.158 | **0.158** | 0.112 |
| **$M_H$-$M_L$** | -116 | -0.278 | -0.291 | -0.077 | 0.233 | **0.265** |
| **$t_{90}$** | -398 | 0.046 | 0.154 | 0.097 | **0.266** | 0.239 |
| **Tensile Strength** | -1168 | -0.221 | -0.317 | **0.085** | 0.012 | 0.072 |
| **Mod 300** | -61 | -0.025 | -1.234 | **-0.007** | -0.013 | -0.058 |
| **EB %** | -783 | -0.064 | -0.216 | -0.043 | **0.064** | -0.122 |
| **Hardness** | -247 | 0.056 | 0.033 | **0.135** | -0.021 | -0.060 |
| **tan δ (0°C)** | -65 | -0.105 | 0.015 | -0.101 | **0.084** | 0.032 |
| **tan δ (60°C)** | -2146 | -0.174 | -0.187 | **-0.036** | -0.182 | -0.148 |
| **E' (30°C)** | -671 | -0.163 | -0.476 | **-0.037** | -0.095 | -0.075 |
| **Average** | -559 | -0.151 | -0.312 | -0.009 | **0.019** | 0.003 |
| **Rank** | 6 | 4 | 5 | 3 | 1 | 2 |

As stated in the table above, leaving one experiment group out leads to lower $R^2$ for all the models and properties compared to LOO. This indicates that none of the models can be used for predicting properties of new experiment. The likely cause in poor performance is the dataset shift. The joint distribution of features and targets in the unseen group does not match the joint distribution in the training samples. The reason can be that many materials were present in only one experiment. These materials most likely include important information about the properties, but the model does not know how to handle it.

## 5.4 Property prediction

KRR performed best on average for LOOCV. When looking at the individual results, not all properties received similar metrics. Table 6 summarizes the average $R^2$, NMAE, and NRMSE scores for all models (excluding LR), corresponding ranks, and top score. Some properties performed better and had higher $R^2$ scores and lower error rates than others. There is even some variation in average rank of each metric. The $R^2$ score of tensile strength is ranked sixth, but it has the lowest error scores. $M_H$-$M_L$ has similar rise in rank for error metrics. tan δ (0°C) has the second highest average $R^2$ score but is 7th and 8th in NMAE and NRMSE scores respectively. $M_L$ and $t_{90}$ also have drop in ranks when comparing $R^2$ to NMAE and NRMSE. This kind of variation highlight the need for multiple metrics. Smaller error does not necessarily mean better explainability of the variance in the data.

Table 6 Differences between the metrics for each property. Average score, the rank based on average score, and the top score are shown on the table. $R^2$ scores over 0.8 and NMAE scores under 7 % are highlighted with green. The best value in each column is bolded.

| | R2 | | | NMAE | | | NRMSE | | |
|---|---|---|---|---|---|---|---|---|---|
| **Target** | **Average** | **Rank** | **Top Score** | **Average** | **Rank** | **Top Score** | **Average** | **Rank** | **Top Score** |
| **$M_L$** | 0.804 | 5 | 0.824 | 7.44 % | 8 | 7.02 % | 10.35 % | 7 | 9.81 % |
| **$M_H$** | 0.744 | 7 | 0.771 | 6.68 % | 6 | 6.33 % | 8.64 % | 5 | 8.17 % |
| **$M_H$-$M_L$** | 0.729 | 8 | 0.755 | 6.50 % | 4 | 6.06 % | 8.34 % | 3 | 7.93 % |
| **$t_{90}$** | 0.814 | 3 | 0.824 | 6.61 % | 5 | 6.16 % | 8.84 % | 6 | 8.60 % |
| **Tensile Strength** | 0.790 | 6 | 0.816 | **5.89 %** | **1** | **5.43 %** | **7.50 %** | **1** | **7.02 %** |
| **Mod 300** | 0.279 | 11 | 0.352 | 10.84 % | 11 | 9.82 % | 18.70 % | 11 | 17.75 % |
| **EB %** | 0.687 | 9 | 0.738 | 9.69 % | 10 | 8.42 % | 12.77 % | 10 | 11.72 % |
| **Hardness** | **0.852** | **1** | **0.883** | 6.04 % | 2 | 5.58 % | 7.97 % | 2 | 7.11 % |
| **tan δ (0°C)** | 0.823 | 2 | 0.852 | 7.40 % | 7 | 7.02 % | 10.39 % | 8 | 9.52 % |
| **tan δ (60°C)** | 0.810 | 4 | 0.852 | 6.45 % | 3 | 5.63 % | 8.64 % | 4 | 7.65 % |
| **E' (30°C)** | 0.552 | 10 | 0.624 | 9.67 % | 9 | 9.05 % | 12.34 % | 9 | 11.34 % |

Table 6 shows that all NMAE scores are lower than NRMSE. NRMSE penalizes large errors more than NMAE. This indicates that on average errors are not that big but there are few large errors in each model that makes NRMSE higher.

One possible explanation in rank changes between $R^2$ and NMAE and NMRSE can be found when looking at the distributions shown in Figure 10. The distribution of tan δ (0°C) is bimodal which can lead to bigger errors. Bimodality suggests that there are two types of data, and this division can be difficult for the model to learn. However, hardness has bimodal distribution too, but the errors are relatively low. $M_L$ and $t_{90}$ on the other hand have skewed distributions. Skewed distribution can lead to model bias towards more frequent values, and the error is smaller even if the model does not capture the variance in data so well. Tensile strength and $M_H$-$M_L$ have more centered distributions with higher peak which can make the errors smaller even when the model is not so accurate.

Other explanation can be also found by looking at the CV values of the distributions. As $R^2$ score is calculated based on the total variance of the target property, it is more difficult for the model to reach high $R^2$ scores for the properties with higher CV. This could explain the difference between hardness and tensile strength as hardness has the lowest CV among the targets. Tensile strength has larger variation than hardness which leads to lower $R^2$ score even when the NMAE and NRMSE are the smallest among the properties. On the other hand, this can explain why modulus 300 is difficult for the models to predict as it has the highest CV among the targets.

Interestingly, tan $\delta$ (0 °C) has quite high CV but has the second highest $R^2$ score. However, the errors are quite large. As shown later in Figure 14, the model for tan $\delta$ (0 °C) estimates lower values relatively well but errors are larger on higher values.

### 5.4.1 Vulcanization characteristics

Vulcanization characteristics were closest to each other of property types in metrics where they ranked in the middle. Vulcanization characteristics are measured with the MDR so the ease of prediction should be consistent. However, different principles are behind the properties like curing time, stiffness or processability, which makes the properties not correlating. $t_{90}$ had the highest $R^2$ score whereas $M_H$-$M_L$ had slightly lower $R^2$ score among vulcanization characteristics. On the other hand, $M_L$ had the highest errors while the rest had more similar errors. $M_L$ had the largest errors even though it had the second best $R^2$ score.

### 5.4.2 Mechanical properties

Hardness and tensile strength were the easiest to predict among mechanical properties. Hardness is measured separately from other mechanical properties, and the measurement is typically the most non-sensitive measurement. Roughness of hardness measuring could lead to a lot of noise, but in this case, it was one of the best performing properties. The reason might be that hardness can be easily controlled with material selection e.g. using softeners, which might explain why it is easy to predict based on the recipe.

Tensile strength, elongation at break %, and modulus 300 are measured together in tensile testing but they had the widest variation in ranking. The differences in prediction precisions are likely due to some other factors than used materials and equipment. The variation can be due to processing of the compound, and the model does not have access to this information, which can lead to lower results. Tensile strength had the smallest errors of all properties whereas modulus 300 had the largest. Figure 12 shows the difference between tensile strength and elongation at break %. For tensile strength, the samples are close to perfect fit except for a single outlier which leads to low error rates. For elongation at break %, the samples are farther away from the perfect fit line.

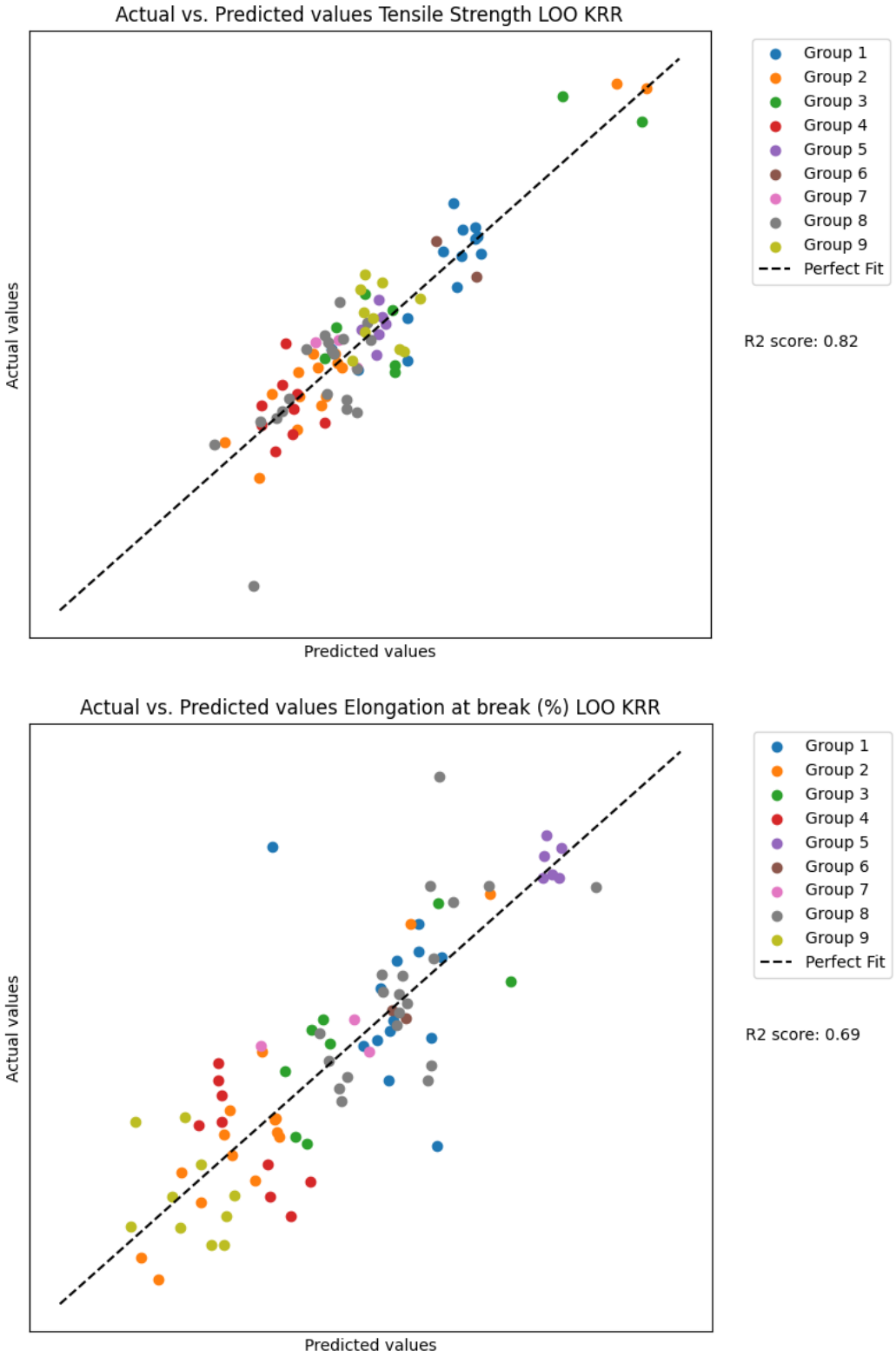


Figure 12 Actual vs. predicted values for tensile strength and elongation at break % using kernel ridge regression for both. In tensile strength the values are more closely packed around perfect fit line whereas in elongation at break % are more spread.

Modulus 300 was the most difficult among all the metrics. This is in line with earlier studies [27]. Modulus 300 is the least straight forward to measure, as the sample can break before reaching the elongation of 300 %. Figure 13 shows that it is likely the reason why the models perform poorly on modulus 300 as they have difficulties to predict the breaking of the sample.

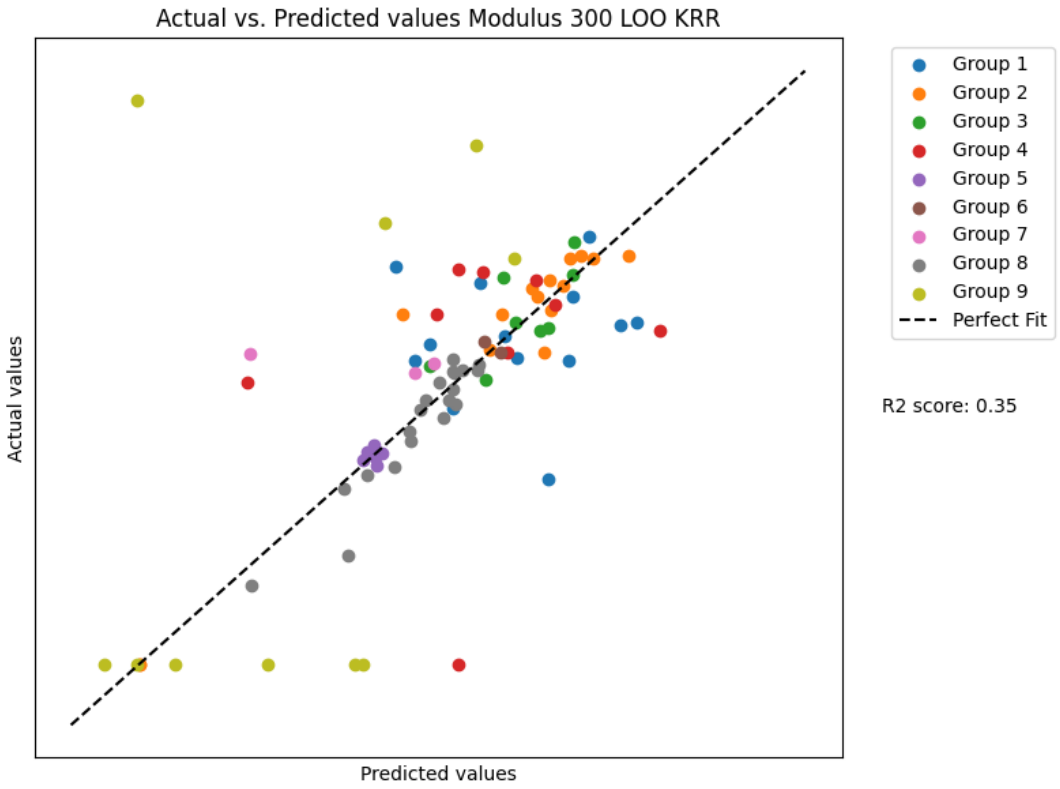


Figure 13 Actual vs. predicted values for modulus 300 modelled using kernel ridge regression.

### 5.4.3 Dynamic mechanical properties

The sample size for dynamic mechanical properties was 66 samples. Tan δ for 0 °C and 60 °C were easy to predict for all models with tan delta (60 °C) being generally easier. E' 30° was more difficult for all the models. The reason can be that while they are related, E' is the storage modulus and tan delta describes the ratio between the loss modulus and storage modulus which are related to different behavior in the compound.

Figure 14 shows the difference between tan δ for 0 °C and 60 °C.

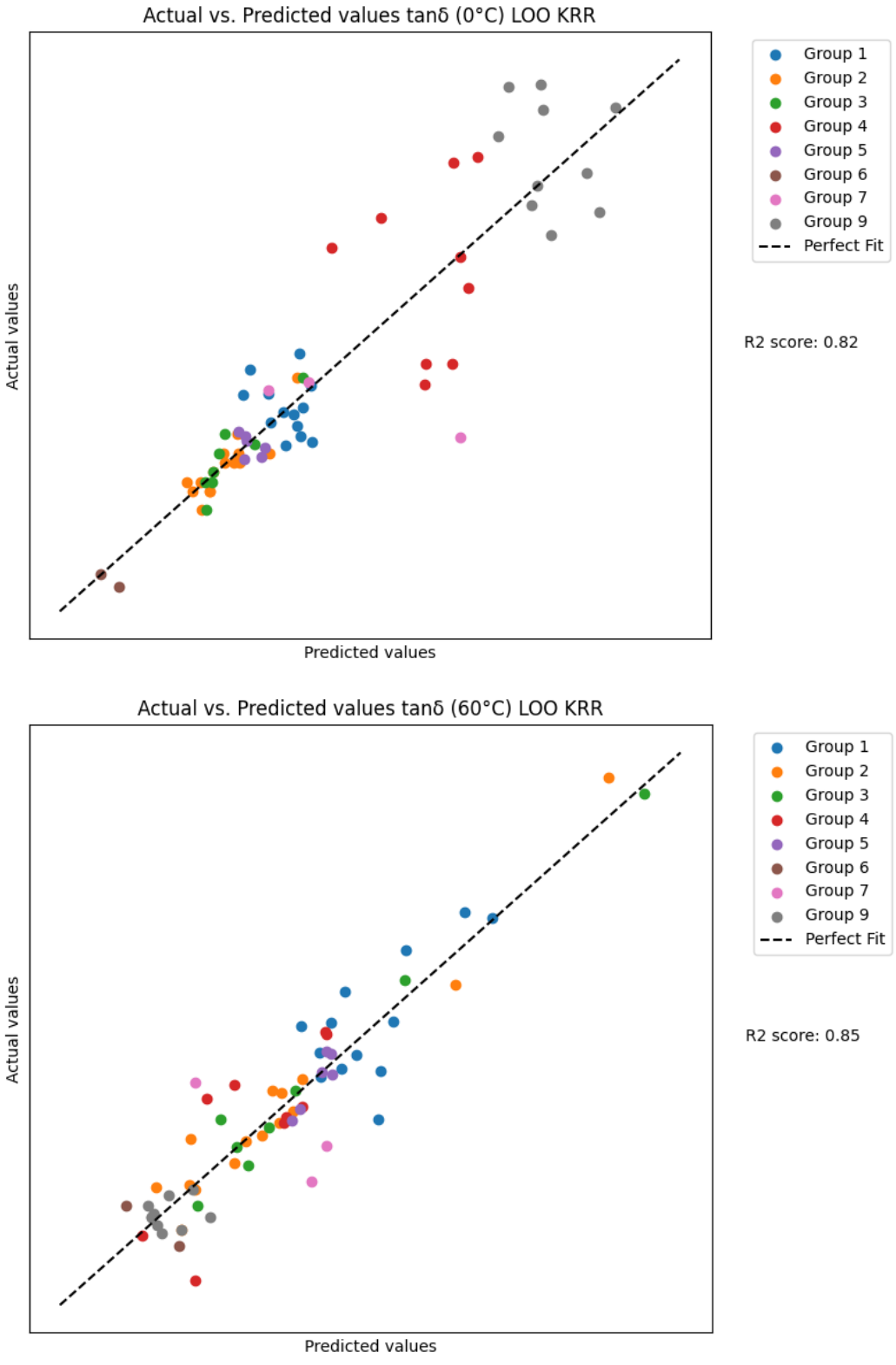


Figure 14 Actual vs. predicted values for tan δ for 0 °C and 60 °C modelled using kernel ridge regression. In 0 °C the groups are more separated, and some groups are visibly closer to the perfect fit line than others. In 60 °the values and experiment groups are spread more evenly along the perfect fit line.

## 5.5 Filtering

The model performance can be improved by filtering some groups e.g. using a subset of the data. It is generally considered good practice to filter lower quality data from the model. However, defining lower quality is not always easy and especially when there is a small dataset, it can be risky. Here some groups are left out in order to boost modeling precision. All the modeling steps were similar to previous models. There was only less data to train with. The best performing model KRR is used here for demonstration.

Filtering is examined by observing the behavior of tensile strength and $M_L$ results using a subset of data. These properties were chosen as arbitrary examples. The results are summarized in table 7. In the full dataset, tensile strength had lowest error while having medium $R^2$ score. $M_L$ had better rank in $R^2$ than in errors. The number of samples was selected into filtered samples based on RMSE analysis where each group was left out and the effect on RMSE score was studied. The final subset was manually confirmed by comparing the metrics to find the best subset. From table 7 we see that all the metrics improved when taking a subset of the data. The difference was bigger for $M_L$ which had lower metrics.

Table 7 Comparison of subset and all data metrics for $M_L$ and tensile strength using KRR and LOOCV.

| Target | $R^2$ | NRMSE | NMAE | Kendall's $\tau$ | p-value ($\tau$) | Number of samples |
|---|---|---|---|---|---|---|
| **$M_L$ Filtered** | 0.895 | 7.35 % | 5.58 % | 0.798 | 3.23E-16 | 51 |
| **$M_L$ all** | 0.796 | 10.56 % | 7.20 % | 0.726 | 7.74E-23 | 86 |
| **Tensile Strength Filtered** | 0.906 | 5.48 % | 4.23 % | 0.726 | 1.01E-14 | 56 |
| **Tensile Strength all** | 0.816 | 7.02 % | 5.43 % | 0.653 | 6.39E-19 | 86 |

Figure 15 shows support for the findings in metrics as all the data points are close to the perfect fit line.

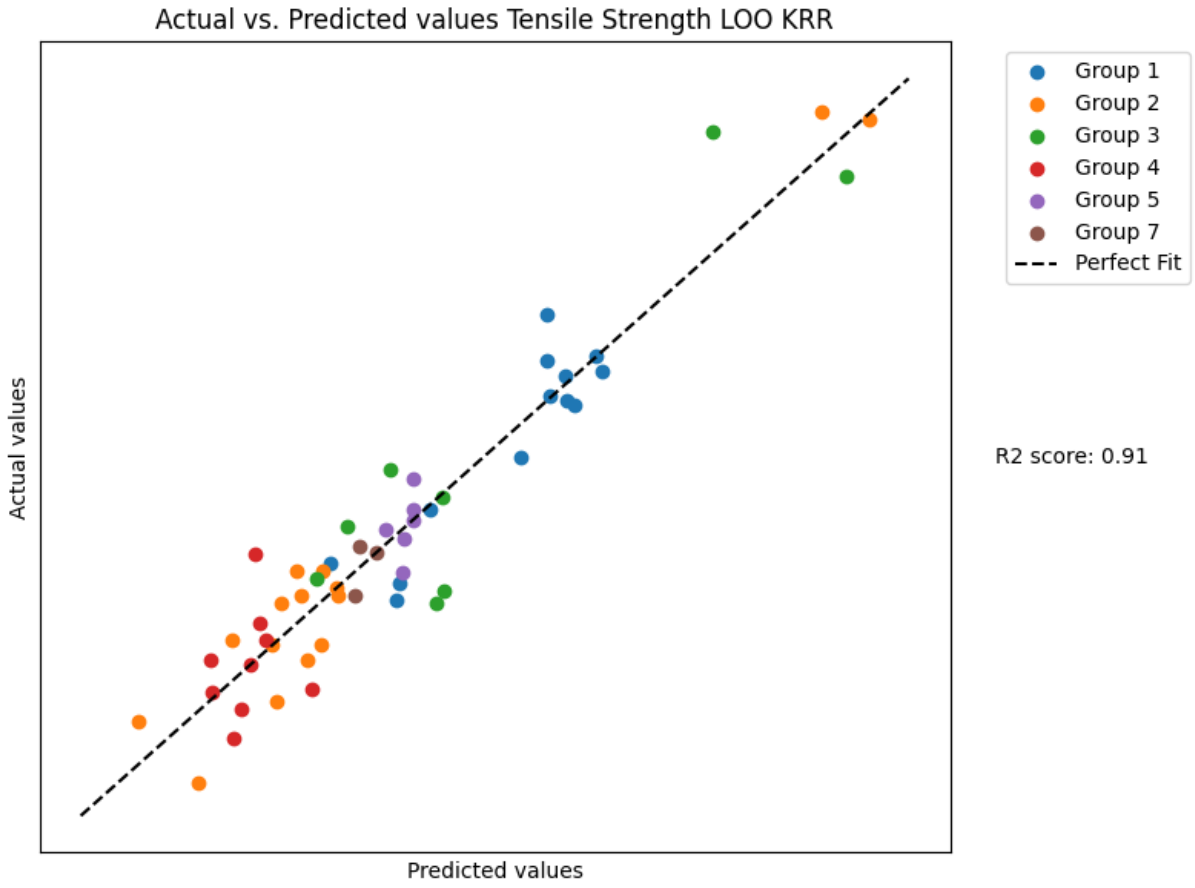


Figure 15 Actual vs. predicted values of tensile strength using a subset of data.

Results could be improved for LOGOCV too. $t_{90}$ modeled with RF received the highest $R^2$ score among the LOGOCV models. It is used here as arbitrary example. The metrics comparison is shown in Table 8. All the metrics improved with data selection.

Table 8 Comparison of subset and all data metrics for $t_{90}$ using random forest and LOGOCV.

| Target | $R^2$ | NRMSE | NMAE | Kendall's $\tau$ | p-value ($\tau$) | Number of samples |
|---|---|---|---|---|---|---|
| **$t_{90}$ Filtered** | 0.691 | 12.06 % | 9.45 % | 0.643 | 2.82E-14 | 66 |
| **$t_{90}$ all** | 0.266 | 17.58 % | 13.64 % | 0.418 | 2.22E-08 | 86 |

The metrics look promising, but Figure 16 tells a different story. The problems are the same as in Figure 11 where the groups are formed vertically and not along the perfect fit line. Despite the improved metrics, this model does not predict the data in a way that could be useful in applications.

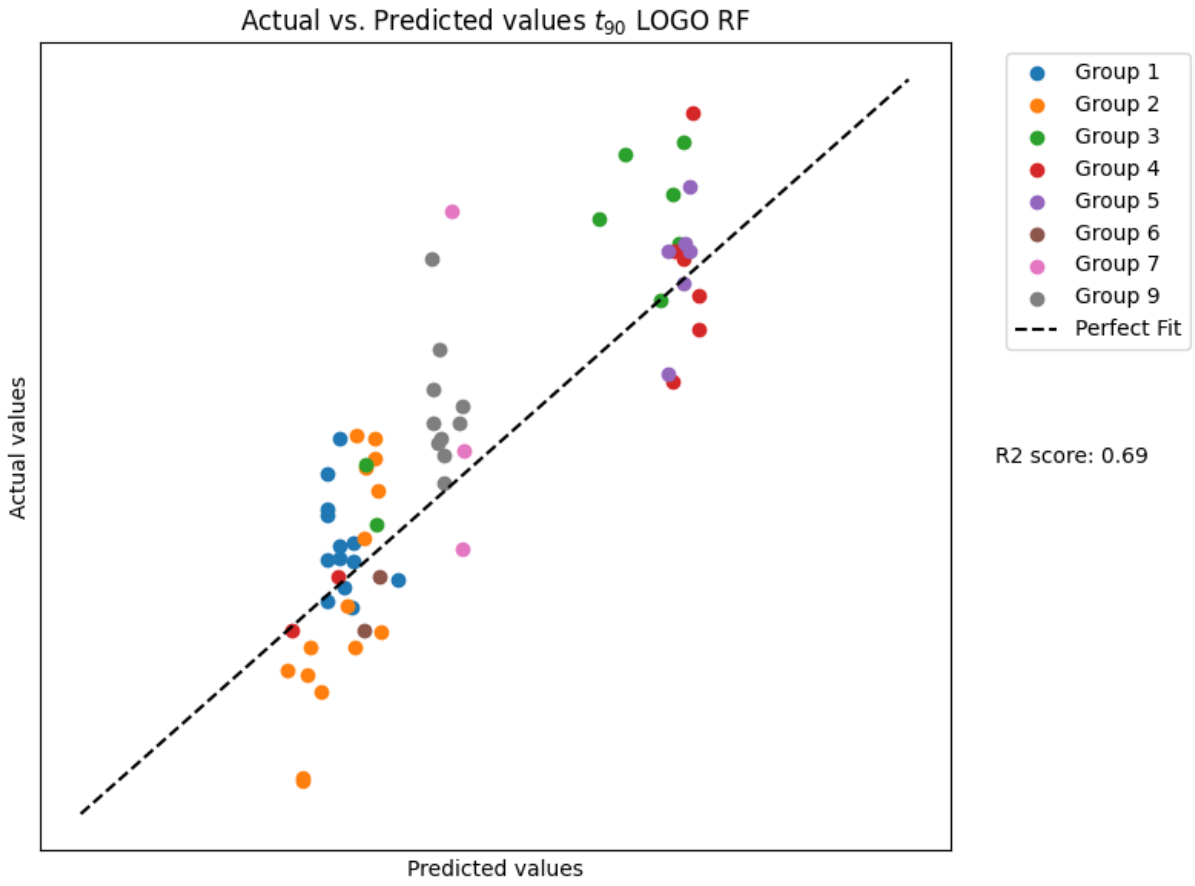


Figure 16 Actual vs. predicted values of $t_{90}$ using subset of data modelled using RF and LOGOCV.

Selecting the data carefully can be beneficial when accuracy is important. This makes the predictions more precise. However, using a subset of the data makes the application range narrower as the model is only applicable in its context and limitations. Using a subset can be thus useful when there needs to be more precise estimation in the limits of already studied samples. Downside is that it makes the model's generalizability worse and some important information is left out of the model, especially on small datasets.

# 6 Discussion and Recommendations

This study confirmed that machine learning methods can be used to predict multiple tire material properties based on their formulation. In total, eleven different properties were studied in this thesis. There was some variation between the properties and how precise the models were.

The analysis of the results revealed several key findings. Most notably, there was no obvious all-round winner among the models. On average, kernel ridge regression was the strongest when evaluated with LOOCV and random forest with LOGOCV respectively. Linear regression performed very poorly on the data but interestingly ridge regression that is similar to linear regression but with the shrinkage parameter performed close to more complex models.

The results suggest that there is variation in the performance of properties measured with same machine which is the case for rheometric properties, tensile stress-strain, and dynamic mechanical properties. The same phenomena has been observed in previous studies. It highlights the special nature of elastomer compounds where relationships between the compounds and properties are not linear. A possible explanation for this finding is that the principles behind each property differ a lot even when they are measured with same device. Most studies have focused on only few metrics even when multiple metrics are used in the tire material development process. This study shows how difficult it is to find one model that can answer real-life problems where multiple vastly different properties are included.

LOOCV gave more optimistic results than LOGOCV. This indicates that there must be enough similar compounds in the dataset which means that the models cannot generalize for completely new compounds with new materials. The size of the dataset and the distribution of the properties likely affects the generalization, and it could be improved by introducing more data to the models.

Using machine learning in predicting rubber compound properties has been studied a lot in recent years and the interest in industry has grown to using machine learning as a tool in research and development. While previous studies have focused on using neural networks, this study excluded them from the analysis due to data limitations. One challenge in comparison is the lack of consistent metrics to compare models from previous studies. Relative error and prediction accuracy have been used in the studies before, but they were not

suitable for the dataset in this study. Nevertheless, relative error can be cautiously compared to normalized errors even if they are not fully comparable. NMAE and NRMSE values in this study were more comparable to older studies in scale but newer studies had significantly lower error rates. $R^2$ is commonly used metric, and many studies have shown very high (>0.95) values while this thesis has lower $R^2$ values that is explainable by the CV values of the properties. The difference is likely caused by the small dataset that was combined from multiple experiments. Furthermore, the proportion of the materials compared to samples was much larger in this thesis.

This thesis included another dimensionless metric besides $R^2$. Kendall's $\tau$ was used to evaluate how well the models can keep the original ranking between the samples. This is important knowledge for practical applications when developing compounds that need to have higher or lower values than other compounds. High $\tau$ value would indicate that the model can determine accurately if the sample is located higher or lower on given property.

## 6.1 Limitations

The biggest limitations of this thesis are related to the dataset. The sample size is small which limits the generalizability. Nevertheless, many materials were covered which gives a broader view to different compound recipes. The samples consists of experiments from different laboratories which introduces naturally some noise in the data which can decrease the performance compared to situation where all the data comes from same experiment and laboratory. Many factors, like human errors or equipment from different manufacturer, can create noise in laboratory experiments. Despite the challenges, there are trends that can be predicted with machine learning in the current dataset.

The lack of universal metrics in the field makes it more difficult to compare models with related studies. The models built in this thesis have quite large errors. This limits the application of the models in the industry problems. NMAE values were between 5-11 % in LOOCV and 12-22 % in LOGOCV for the best models. LOGOCV evaluation showed how poorly the models can generalize outside the studied experiments. This makes studying new materials without prior experiments very difficult. This is a common problem in machine learning as discussed by Moreno-Torres et al.[34]

There are a lot of production related factors that affect the studied properties. They can have a larger role in some properties that are used to measure production consistency. Processing

conditions can change the results, but they were not studied in this thesis. These factors can affect more the properties where prediction precision was weaker.

Despite these limitations, the study has application in practice. The error marginal is too wide when making decisions but can be acceptable when designing the recipes, looking for alternatives, or planning the experiments. In addition, it is possible to get preliminary results even for new compounds that do not have similar compounds in the dataset.

## 6.2 Recommendations and future research

Environmental issues require new materials to be used in tires to replace harmful ones like 6-PPD. However, it is difficult to replace a material in tire recipe and keep the same properties. It can take a lot of time and trials to find a suitable substitute. Machine learning has potential to aid in the task but that requires data and high accuracy. In the future, it would be useful to study how much and what kind of data is needed to achieve high prediction precision and very low errors while having a large variety of materials, so that models' prediction only need to be verified in the laboratory. There are hundreds of different materials and even more grades for material engineer to choose from and testing all the quantity combination is not possible. Important question is how much data will be enough. Creating samples in the laboratory takes time and money. Including data from the processing is reasonable for some properties. However, this will lead to infinitely large dataset. A feasible option is to map what factors and materials are the most crucial for each property and make customized datasets with less features for each one to get the highest predictability.

Less studied opportunity for machine learning is to apply it into reverse engineering purposes where the materials of tire piece are predicted based on the laboratory results. This will require different dataset as tests used in reverse engineering are different for compound development.

Kendall's $\tau$ is used in this study as a dimensionless metric to evaluate if the models can rank target variable correctly. The use of $\tau$ needs to be further studied and verified in research to see whether it can provide a new helpful tool to evaluate models' performance. In the future this could be used as another dimensionless metric to compare model performance between different studies.

# 7 Conclusions

This thesis studies if multiple tire compound properties can be predicted based on the material composition. The thesis employes several machine learning methods, and the prediction precision is evaluated through multiple metrics. Based on the results, machine learning can predict properties, but error rates limit applicability. The analysis reveals that there is no all-round best model. KRR has the highest score when LOOCV is used and RF when LOGOCV is used. LOOCV reaches much better results than LOGOCV which indicates that the models cannot extrapolate to unseen materials.

There are differences in prediction precision between the properties. Some of the properties have higher prediction precision on the metrics, e.g. tensile strength and hardness. Some other properties, such as modulus 300 and E' (30°C) are clearly more difficult to predict. There are multiple possible reasons why some models and properties perform better than others. Data distribution can affect the model's performance. There are different principles and relationships behind each property that influences the model complexity and data point consistency. It is also likely that other factors than materials have important roles for some properties.

This study compares properties on different scales. Consequently, dimensionless metrics are needed for comparison. $R^2$ is commonly used metric in regression tasks. In addition, the thesis implements NRMSE and NMAE and introduced the use of Kendall's $\tau$ as dimensionless metric for datasets where the order is important. These dimensionless metrics make future comparisons of different studies easier.

Machine learning can be helpful tool in tire material engineering. It shows promise in designing compositions and planning experiments when the engineer can estimate the results roughly beforehand. By analyzing the data and its distribution, it is possible to reveal what kind of data is needed and which data regions are underrepresented. Then resources can be focused on collecting the most valuable missing data. Especially when trying to find new sustainable materials, analyzing the data and predicting the results can provide a useful tool to reduce the amount of overall laboratory experiments needed.

# References


[1] B.E. Lindenmuth, An overview of tire technology, The Pneumatic Tire 1 (2006) 13136.

[2] F. Sommer, V. Dietze, A. Baum, J. Sauer, S. Gilge, C. Maschowski, R. Gieré, Tire Abrasion as a Major Source of Microplastics in the Environment, Aerosol Air Qual. Res. 18 (2018) 2014–2028. https://doi.org/10.4209/aaqr.2018.03.0099.

[3] J.S. Dick, 1. Rubber Compounding: Introduction, Definitions, and Available Resources, in: Rubber Technology - Compounding and Testing for Performance, 3rd ed., Hanser Publications, Munich, Germany, 2020.

[4] M. Manges, How Are Tiremakers, Suppliers Using AI?, Modern Tire Dealer (2023). https://www.moderntiredealer.com/suppliers/article/33009732/artificial-intelligence-how-are-tiremakers-suppliers-using-it (accessed March 18, 2025).

[5] Y. Huang, Q. Chen, Z. Zhang, K. Gao, A. Hu, Y. Dong, J. Liu, L. Cui, A Machine Learning Framework to Predict the Tensile Stress of Natural Rubber: Based on Molecular Dynamics Simulation Data, Polymers 14 (2022) 1897. https://doi.org/10.3390/polym14091897.

[6] S. Pang, J. Luo, Y. Wu, Properties Prediction and Design of Tire Tread Composites Using Machine Learning, Macromolecular Theory and Simulations 29 (2020) 1900063. https://doi.org/10.1002/mats.201900063.

[7] B. Rodgers, W. Waddell, Chapter 14 - Tire Engineering, in: J.E. Mark, B. Erman, C.M. Roland (Eds.), The Science and Technology of Rubber (Fourth Edition), Fourth Edition, Academic Press, Boston, 2013: pp. 653–695. https://doi.org/10.1016/B978-0-12-394584-6.00014-5.

[8] B. Rodgers, Natural Rubber and Other Naturally Occurring Compounding Materials, in: Rubber Compounding Chemistry and Applications, 2nd ed., CRC Press, Boca Raton, FL, 2016: p. 624.

[9] S.D. Gupta, R. Mukhopadhyay, K.C. Baranwal, A.K. Bhowmick, Reverse engineering of rubber products: Concepts, tools, and techniques, CRC press, 2013.

[10] Hankook Tire & Technology, 2025 1Q Financial Results, (2025). https://www.hankooktire.com/wsvc/api/pdf-viewer/document.servlet?documentPath=/content/dam/hankooktire/common/ir/GO/en/2024/04/24_1Q_IR_ENG_F.pdf (accessed April 6, 2025).

[11] B. Rodgers, W. Waddell, Chapter 9 - The Science of Rubber Compounding, in: J.E. Mark, B. Erman, C.M. Roland (Eds.), The Science and Technology of Rubber (Fourth Edition), Fourth Edition, Academic Press, Boston, 2013: pp. 653–695. https://doi.org/10.1016/B978-0-12-394584-6.00014-5.

[12] H.-W. Engels, H.-J. Weidenhaupt, M. Pieroth, W. Hofmann, K.-H. Menting, T. Mergenhagen, R. Schmoll, S. Uhrlandt, Rubber, 9. Chemicals and Additives, in: Ullmann's Encyclopedia of Industrial Chemistry, John Wiley & Sons, Ltd, 2011. https://doi.org/10.1002/14356007.a23_365.pub3.

[13] N. Hewitt, Compounding Precipitated Silica in Elastomers: Theory and Practice, 1st ed., Elsevier Science & Technology Books, Chantilly, 2007.

[14] A.Y. Coran, Chapter 7 - Vulcanization, in: J.E. Mark, B. Erman, C.M. Roland (Eds.), The Science and Technology of Rubber (Fourth Edition), Fourth Edition, Academic Press, Boston, 2013: pp. 653–695. https://doi.org/10.1016/B978-0-12-394584-6.00014-5.

[15] B.H. To, 15. Sulfur Cure Systems, in: Rubber Technology - Compounding and Testing for Performance, 3rd ed., Hanser Publications, Munich, Germany, 2020.

[16] B. Seiwert, M. Nihemaiti, M. Troussier, S. Weyrauch, T. Reemtsma, Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater, Water Research 212 (2022) 118122. https://doi.org/10.1016/j.watres.2022.118122.

[17] N. Hewitt, P.A. Ciullo, Natural Rubber, Polyisoprene, in: Rubber Formulary, William Andrew Publishing/Plastics Design Library, 1999: p. 9.

[18] J.S. Dick, 3. Vulcanizate Physical Properties, Performance Characteristics, and Testing, in: Rubber Technology - Compounding and Testing for Performance, 3rd ed., Hanser Publications, Munich, Germany, 2020.

[19] R. Ding, A.I. Leonov, A kinetic model for sulfur accelerated vulcanization of a natural rubber compound, Journal of Applied Polymer Science 61 (1996) 455–463. https://doi.org/10.1002/(SICI)1097-4628(19960718)61:3<455::AID-APP8>3.0.CO;2-H.
[20] J.S. Dick, Curemeters, in: Practical Rubber Rheology and Dynamic Properties, Hanser Publishers, Germany, 2023: pp. 1–1.
[21] P. Konečnỳ, M. Černỳ, J. Voldanova, J. Maláč, J. Šimoník, Dynamic mechanical properties of filled styrene butadiene rubber compounds: comparison of tensile and shear data, Polymers for Advanced Technologies 18 (2007) 122–127.
[22] E. Terrill, A. Bashir, R. Meser, Understanding variables impacting tire traction, Rubber & Plastics News (2019) 14, 16–18.
[23] D.J. Zanzig, P.H. Sandstrom, M.J. Crawford, J.J.A. Verthe, C.A. Losey, Tire with silica reinforced tread, US5614580A, 1997. https://patents.google.com/patent/US5614580A/en?oq=5614580 (accessed May 15, 2025).
[24] G. Heeps, Feature: Digital updates, Tire Technology International (2023). https://www.tiretechnologyinternational.com/features/digital-updates.html (accessed March 18, 2025).
[25] G.A. Schwartz, Prediction of rheometric properties of compounds by using artificial neural networks, Rubber Chemistry and Technology 74 (2001) 116–123.
[26] P. Xu, H. Chen, M. Li, W. Lu, New opportunity: machine learning for polymer materials design and discovery, Advanced Theory and Simulations 5 (2022) 2100565.
[27] V. Vijayabaskar, R. Gupta, P. Chakrabarti, A.K. Bhowmick, Prediction of properties of rubber by using artificial neural networks, Journal of Applied Polymer Science 100 (2006) 2227–2237.
[28] B. Wang, J.H. Ma, Y.P. Wu, Application of artificial neural network in prediction of abrasion of rubber composites, Materials & Design 49 (2013) 802–807.
[29] Z. Uruk, A. Kiraz, V. Deniz, A comparison of machine learning methods to predict rheometric properties of rubber compounds, J Rubber Res 25 (2022) 265–277. https://doi.org/10.1007/s42464-022-00170-7.
[30] A.J. Román, S. Qin, J.C. Rodríguez, L.D. González, V.M. Zavala, T.A. Osswald, Natural Rubber Blend Optimization via Data-Driven Modeling: The Implementation for Reverse Engineering, Polymers 14 (2022) 2262-.
[31] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, Peerj Computer Science 7 (2021) e623.
[32] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, Npj Computational Materials 4 (2018) 25.
[33] C.M. Bishop, Pattern recognition and machine learning, Springer New York, 2006.
[34] J.G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N.V. Chawla, F. Herrera, A unifying view on dataset shift in classification, Pattern Recognition 45 (2012) 521–530.
[35] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, An introduction to statistical learning with applications in python, (2023).
[36] C. Tofallis, A better measure of relative prediction accuracy for model selection and model estimation, Journal of the Operational Research Society 66 (2015) 1352–1362.
[37] kendalltau — SciPy v1.16.0 Manual, (n.d.). https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html (accessed June 30, 2025).
[38] D.C. Montgomery, E.A. Peck, G.Geoffrey. Vining, Introduction to Linear Regression Analysis., 5th ed., John Wiley & Sons, Incorporated, Hoboken, 2012.
[39] A.E. Hoerl, R.W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics 12 (1970) 55–67.
[40] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (2004) 199–222.
[41] T. Zou, Y. Dou, H. Mi, J. Zou, Y. Ren, Support vector regression for determination of component of compound oxytetracycline powder on near-infrared spectroscopy, Analytical Biochemistry 355 (2006) 1–7.
[42] V.N. Vapnik, The Nature of Statistical Learning Theory, 1st ed., Springer, New York, NY, 1998.

[43] V. Vovk, Kernel ridge regression, in: Empirical Inference: Festschrift in Honor of Vladimir n. Vapnik, Springer, 2013: pp. 105–116.
[44] scikit-learn developers, Pairwise metrics, affinities and kernels — Laplacian kernel, (2025). https://scikit-learn.org/stable/modules/metrics.html#laplacian-kernel (accessed October 2, 2025).
[45] L. Breiman, Random forests, Machine Learning 45 (2001) 5–32.
[46] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: Data mining, inference, and prediction. 2nd ed., corrected at 11th printing, 2nd ed., Springer, New York, 2016.
[47] scikit-learn developers, StandardScaler, Scikit-Learn 1.7.1 Documentation (2025). https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#sklearn.preprocessing.StandardScaler (accessed August 27, 2025).
[48] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A practical guide to support vector classification, (2003).

# Appendices

## Appendix 1 LOOCV NMAE, NRMSE, and $\tau$

Each appendix is numbered and given a heading.

| | LOO NMAE | | | | | |
|---|---|---|---|---|---|---|
| **Target** | **LR** | **RR** | **KRR** | **SVR** | **RF** | **GBR** |
| **$M_L$** | 59.30 % | 7.70 % | 7.20 % | **7.02 %** | 7.75 % | 7.53 % |
| **$M_H$** | 80.03 % | 6.95 % | **6.33 %** | 6.62 % | 6.70 % | 6.79 % |
| **$M_H$-$M_L$** | 76.93 % | 6.97 % | **6.06 %** | 6.80 % | 6.34 % | 6.36 % |
| **$t_{90}$** | 124.54 % | 6.81 % | **6.16 %** | 6.88 % | 6.75 % | 6.42 % |
| **Tensile Strength** | 97.73 % | 6.05 % | **5.43 %** | 6.27 % | 5.86 % | 5.81 % |
| **Mod 300** | 63.19 % | 12.50 % | **9.82 %** | 10.35 % | 11.42 % | 10.12 % |
| **EB %** | 144.03 % | 11.00 % | 9.45 % | 10.69 % | 8.90 % | **8.42 %** |
| **Hardness** | 112.17 % | **5.58 %** | 5.58 % | 5.58 % | 6.75 % | 6.71 % |
| **tan δ (0°C)** | 30.70 % | 7.70 % | **7.02 %** | 8.18 % | 7.04 % | 7.04 % |
| **tan δ (60°C)** | 63.51 % | 6.60 % | **5.63 %** | 6.92 % | 7.18 % | 5.93 % |
| **E' (30°C)** | 88.98 % | 9.16 % | **9.05 %** | 9.63 % | 10.46 % | 10.04 % |
| **Average** | 85.56 % | 7.91 % | **7.07 %** | 7.72 % | 7.74 % | 7.38 % |
| **Rank** | 6 | 5 | **1** | 3 | 4 | 2 |
| | **LOO NRMSE** | | | | | |
| **Target** | **LR** | **RR** | **KRR** | **SVR** | **RF** | **GBR** |
| **$M_L$** | 117.48 % | 11.12 % | 10.56 % | 10.28 % | 9.95 % | **9.81 %** |
| **$M_H$** | 232.56 % | 8.98 % | **8.17 %** | 8.78 % | 8.75 % | 8.54 % |
| **$M_H$-$M_L$** | 228.44 % | 8.89 % | **7.93 %** | 8.68 % | 8.18 % | 7.99 % |
| **$t_{90}$** | 242.57 % | 8.93 % | **8.60 %** | 8.70 % | 9.09 % | 8.87 % |
| **Tensile Strength** | 206.70 % | 7.38 % | **7.02 %** | 7.79 % | 7.70 % | 7.61 % |
| **Mod 300** | 188.91 % | 20.05 % | 17.77 % | **17.75 %** | 19.26 % | 18.68 % |
| **EB %** | 335.97 % | 13.80 % | 12.76 % | 13.70 % | 11.86 % | **11.72 %** |
| **Hardness** | 276.67 % | 7.35 % | 7.33 % | **7.11 %** | 9.00 % | 9.07 % |
| **tan δ (0°C)** | 56.99 % | 11.18 % | 10.42 % | 10.73 % | **9.52 %** | 10.11 % |
| **tan δ (60°C)** | 139.73 % | 9.39 % | **7.65 %** | 8.92 % | 9.49 % | 7.75 % |
| **E' (30°C)** | 209.61 % | 11.41 % | **11.34 %** | 12.09 % | 14.04 % | 12.82 % |
| **Average** | 203.24 % | 10.77 % | **9.96 %** | 10.41 % | 10.62 % | 10.27 % |
| **Rank** | 6 | 5 | **1** | 3 | 4 | 2 |
| | **LOO $\tau$** | | | | | |

| Target | LR | RR | KRR | SVR | RF | GBR |
|---|---|---|---|---|---|---|
| $M_L$ | 0.355 | 0.744 | 0.726 | **0.774** | 0.714 | 0.730 |
| $M_H$ | 0.333 | 0.651 | 0.638 | **0.660** | 0.620 | 0.609 |
| $M_H$-$M_L$ | 0.338 | 0.618 | **0.625** | 0.583 | 0.566 | 0.577 |
| $t_{90}$ | 0.163 | 0.713 | 0.735 | 0.715 | 0.728 | 0.741 |
| Tensile Strength | 0.227 | 0.606 | **0.653** | 0.599 | 0.632 | 0.634 |
| Mod 300 | 0.324 | 0.438 | **0.541** | 0.511 | 0.471 | 0.527 |
| EB % | 0.262 | 0.600 | 0.656 | 0.598 | 0.669 | 0.680 |
| Hardness | 0.220 | 0.729 | **0.732** | 0.708 | 0.589 | 0.595 |
| tan δ (0°C) | 0.317 | 0.705 | **0.760** | 0.645 | 0.744 | 0.758 |
| tan δ (60°C) | 0.295 | 0.683 | **0.714** | 0.687 | 0.680 | 0.705 |
| E' (30°C) | 0.143 | 0.460 | **0.467** | 0.394 | 0.326 | 0.421 |
| Average | 0.271 | 0.632 | **0.659** | 0.625 | 0.613 | 0.634 |
| Rank | 6 | 3 | **1** | 4 | 5 | 2 |

## Appendix 2 LOGOCV NMAE, NRMSE, and $\tau$

| | LOGO NMAE | | | | | |
|---|---|---|---|---|---|---|
| Target | LR | RR | KRR | SVR | RF | GBR |
| $M_L$ | 287.22 % | 23.74 % | 22.00 % | **18.13 %** | 21.58 % | 21.98 % |
| $M_H$ | 167.88 % | 15.84 % | 16.71 % | 14.35 % | **11.49 %** | 11.94 % |
| $M_H$-$M_L$ | 126.01 % | 14.58 % | 14.79 % | 12.56 % | **10.02 %** | 10.15 % |
| $t_{90}$ | 279.68 % | 16.41 % | 14.28 % | 15.44 % | 13.64 % | **12.48 %** |
| Tensile Strength | 364.84 % | 13.78 % | 14.39 % | **12.08 %** | 12.92 % | 12.54 % |
| Mod 300 | 151.61 % | 16.35 % | 29.11 % | **15.31 %** | 16.55 % | 17.41 % |
| EB % | 412.46 % | 19.67 % | 20.59 % | 19.54 % | **17.82 %** | 19.62 % |
| Hardness | 269.24 % | 15.84 % | **15.35 %** | 15.69 % | 15.98 % | 16.65 % |
| tan δ (0°C) | 157.28 % | 20.23 % | **18.07 %** | 20.35 % | 18.16 % | 18.30 % |
| tan δ (60°C) | 687.54 % | 16.98 % | 18.09 % | **16.38 %** | 17.87 % | 16.48 % |
| E' (30°C) | 358.95 % | 14.56 % | 16.95 % | **13.92 %** | 14.65 % | 14.29 % |
| Average | 296.61 % | 17.09 % | 18.21 % | 15.79 % | **15.52 %** | 15.62 % |
| Rank | 6 | 4 | 5 | 3 | **1** | 2 |
| | LOGO NRMSE | | | | | |
| Target | LR | RR | KRR | SVR | RF | GBR |

| | | | | | | |
|---|---|---|---|---|---|---|
| **$M_L$** | 374.56 % | 27.89 % | 27.89 % | **22.84 %** | 26.61 % | 25.88 % |
| **$M_H$** | 261.81 % | 19.57 % | 20.84 % | 18.39 % | **15.68 %** | 16.10 % |
| **$M_H$-$M_L$** | 173.19 % | 18.12 % | 18.22 % | 16.64 % | 14.04 % | **13.74 %** |
| **$t_{90}$** | 409.79 % | 20.04 % | 18.87 % | 19.49 % | **17.58 %** | 17.90 % |
| **Tensile Strength** | 559.70 % | 18.09 % | 18.78 % | **15.66 %** | 16.27 % | 15.77 % |
| **Mod 300** | 173.80 % | 22.31 % | 32.95 % | **22.12 %** | 22.19 % | 22.68 % |
| **EB %** | 640.82 % | 23.60 % | 25.24 % | 23.37 % | **22.14 %** | 24.25 % |
| **Hardness** | 327.67 % | 20.23 % | 20.47 % | **19.36 %** | 21.03 % | 21.43 % |
| **tan δ (0°C)** | 200.93 % | 26.00 % | 24.55 % | 25.95 % | **23.68 %** | 24.34 % |
| **tan δ (60°C)** | 921.27 % | 21.54 % | 21.67 % | **20.24 %** | 21.61 % | 21.31 % |
| **E' (30°C)** | 479.55 % | 19.94 % | 22.47 % | **18.84 %** | 19.35 % | 19.17 % |
| **Average** | 411.19 % | 21.58 % | 22.90 % | 20.26 % | **20.02 %** | 20.23 % |
| **Rank** | 6 | 4 | 5 | 3 | **1** | 2 |
| | **LOGO $\tau$** | | | | | |
| **Target** | **LR** | **RR** | **KRR** | **SVR** | **RF** | **GBR** |
| **$M_L$** | **0.308** | -0.363 | -0.124 | 0.005 | -0.108 | -0.579 |
| **$M_H$** | 0.266 | -0.322 | 0.068 | -0.461 | **0.297** | 0.196 |
| **$M_H$-$M_L$** | 0.102 | -0.224 | 0.121 | -0.323 | 0.344 | 0.374 |
| **$t_{90}$** | -0.081 | 0.326 | 0.315 | 0.304 | 0.418 | 0.442 |
| **Tensile Strength** | 0.008 | -0.247 | -0.031 | -0.105 | 0.021 | 0.048 |
| **Mod 300** | 0.165 | 0.027 | **0.250** | 0.106 | 0.027 | -0.291 |
| **EB %** | -0.138 | -0.065 | 0.057 | 0.061 | **0.185** | -0.316 |
| **Hardness** | 0.092 | **0.409** | 0.393 | 0.376 | 0.211 | 0.167 |
| **tan δ (0°C)** | -0.043 | 0.181 | 0.121 | **0.222** | 0.172 | 0.144 |
| **tan δ (60°C)** | **0.443** | -0.360 | -0.207 | -0.280 | -0.382 | -0.404 |
| **E' (30°C)** | 0.031 | -0.149 | 0.049 | -0.278 | -0.001 | 0.050 |
| **Average** | 0.105 | -0.071 | 0.092 | -0.034 | **0.108** | -0.016 |
| **Rank** | 2 | 6 | 3 | 5 | **1** | 4 |