# SUPERVISED LEARNING WITH REGRESSION

Michael Brosnan

CIS 435 PRACTICAL DATA SCIENCE USING MACHINE LEARNING

PROF KAKADE

## Introduction

Advertising plays a crucial role in the way many of the world companies spend and obtain their money. Companies often use marketing as their main way to reach their buying audience. Many companies have great products but ultimately fail because they were not able to market effectively to a segment of consumers that were most likely to use and appreciate their product (*Why is marketing so important* 2019). Marketing is important for several reasons- the first of which is that it informs your customers on the basics of your product and how it can improve their lives. Secondly, smart marketing can even the playing field between large corporations and small mom-and-pop retailers, as an effective marketing campaign can reach directly to consumers and change the way they think about your product. Marketing also helps you sell and grow your business by keeping your current clientele up to date with any changes or improvements to your product and helps to reach new clientele that could help you sell more and grow your profits. As a result, companies are investing large amounts of time, energy, and resources into using data and analytics to streamline their marketing initiative to provide the right marketing to the right consumers at the right time at the most efficient cost.

**Commented [MB1]:** https://movingtargets.com/blog/business-marketing/why-marketing-is-so-important/#:~:text=Marketing%20is%20important%20because%20it%20helps%20you%20sell%20your%20products,is%20ultimately%20what%20drives%20sales.

## Business Problem

This brings us to the business problem at hand, which is to suggest the correct media channels. At a high level, this means to help formulate a media campaign strategy by suggesting media channels which effectively produce the best sales and eliminate those that are not effective in producing sales. This will allow the marketing team to dedicate more resources towards the effective methods while not spending on the ineffective methods, which ultimately is the business objective of the project.

The data will be discussed and explored in more detail later in the paper, but as an overview it outlines the amount of dollars in thousands of dollars spend between TV, radio, newspaper, Google, Facebook, and LinkedIn and the amount of product sold also in the thousands. This will be the source of our analysis, and the goal of that analysis will be to help the marketing department determine effective and ineffective means of advertisement. Success would be to identify and train a model or set of models that will help identify patterns in the data that would indicate if an advertising medium were effective or ineffective at producing sales.

In order to effectively produce the end result mentioned above, certain steps within the processes will need to be adhered to. We will follow the CRISP DM process which includes business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Business understanding was accomplished above when we looked at the business problem, determined the business objective as described, determined goals, and the necessary steps to attain those goals. In the following pages, while accomplishing other tasks along the way, we will look to understand the data through an initial exploration of the data as well as the associated fields and prepare the data by cleaning it constructing it and formatting it in order to be effective in our model. We will then determine which models would work best for the data and the objectives then build and assess that model before evaluating it on its results while also reviewing the process that led to those results. After that comes deployment, which for this project is essential producing the final report and project review, which will offer the marketing team suggestions on how to best move forward to most efficiently market the product (*CRISP-DM* 2021).

**Commented [MB2]:** datascience-pm.com/crisp-dm-2/

Once that process is complete, the project is not then over. The process will then continue as the business problem evolves or the data changes and improves. This will lead to the cycle restarting and the same steps being used to continually improve the models that have been created and to better solve the business problem.

## Machine Learning Applications

Marketing is a key aspect for many businesses, and there are many real world examples of companies using machine learning to get better results from their marketing campaigns. One example is using data to better target ads to certain audiences. Often we think of ads as a one size fits all activity, but if you are able to segment the population using different characteristics, you can better focus your different marketing strategies on specific groups of people. Greenpal is an example of a company that used targeted marketing to produce better business results. They ran ads in the Nashville area and overall they deemed the results of the marketing good, they wanted to improve their click through and conversion rates by making the information more relevant to the viewer. They used census data and K-means clustering to segment different neighborhoods and populations in the Nashville area and noticed that there was a neighborhood that had a more price sensitive demographic. For this area, they changed their marketing headline to “The Cheapest Lan Mowing in Nashville” and saw dramatically better results with the targeted ads compared to the general marketing strategy (*5 killer examples of data-driven marketing* 2021).

**Commented [MB3]:** https://www.wordstream.com/blog/ws/2016/08/25/data-driven-marketing

There are also generic applications of machine learning that can be used to get a better idea of the most effective ways of marketing a product. Forecast targeting, like segment targeting, trigger targeting, and predictive targeting, is a popular and impactful tool using by companies to target specific users, usually ones that are most likely to purchase their product. Forecast targeting predicts the possibility of a user making a purchase in a certain number of days. Predictive targeting is specifically useful when an organization has many parameters and wants to use all the possible combinations of those parameters to make the most informed decisions. Segment and trigger are more commonly used when there are less parameters available. Similar to what we saw with the clustering, forecast targeting can allow you to better target certain populations, and it can focus your resources on consumers that are likely to make a purchase; therefore, increasing your return on investment from the marketing department (Vlada Malysheva, 2021).

**Commented [MB4]:** https://www.owox.com/blog/articles/machine-learning-in-marketing/

A third example of machine learning business applications is helping to understand the customer lifetime value; which, in turn, helps you understand and predict future revenue and business success. Lifetime value helps you determine the profit your company can expect from a customer over the life of their relationship with your company. Taking all these customers into account can help you determine future revenue and how much success the business is likely to have on their current path (Diachuk, 2021). This data and analysis can be used to segment customers into populations that can reduce churn rate and ultimately help determine the effectiveness of advertising campaigns. Breaking down data into lifetime value can help a company determine their per-user advertising budget, which will help determine the effectiveness of an advertising campaign. Often, a logistic regression algorithm can be used to determine lifetime value.

**Commented [MB5]:** https://www.owox.com/blog/use-cases/customer-lifetime-value/

These are just three of the ways in which machine learning has business applications, specifically in the marketing department. Many of these are focused on personalization of ads and targeted products, but it has other use cases as well. Items such as the headlines and copy that are used in marketing can also be focused, even the design format like layout color sets and sizing (Columbus, 2018). These items along with things like price optimization and churn rate can be hugely influential to a business's success; and the use of machine learning can help an organization understand their clientele better and understand the incentives of their consumers.

**Commented [MB6]:** https://www.forbes.com/sites/louiscolumbus/2018/02/25/10-ways-machine-learning-is-revolutionizing-marketing/?sh=3129c0005bb6

## Machine Learning Applications

For the business problem stated and the data provided, I used three regression algorithms to identify what the best medium would be to advertise to produce the most sales and at the most efficient cost. Those three algorithms are Linear Regression, Least Absolute Shrinkage Selector Operator (Lasso) Regression, and Ridge Regression. Each of these algorithms offer different strengths and weaknesses, which we will discuss in this section, and we will explore their performance on this dataset through further analysis.

I started out with a simple linear regression model, which "fits a linear model with coefficients w = (w1, …, wp) to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation (*Sklearn.linear_model.linearregression*)." One of the biggest advantages of the linear regression model is its simplicity and the ease at which it can be implemented. This is similar to the other algorithms we used, however, this model was up and running with just a few lines of code, after all the cleaning and preparation. It is also best to use when you know the independent and dependent variables have a linear relationship because it is not as complicated to understand and implement as other algorithms (*ML - advantages and disadvantages of linear regression* 2020).

**Commented [MB7]:** https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html?highlight=linear%20regression#sklearn.linear_model.LinearRegression

While this model is easy to implement, it does have its downsides as well. One of the larger disadvantages of the linear regression model is that it can be massively affected by outliers- which can have a huge negative effect on the model that is being created. Another issue related to the simplicity of the algorithm, is assuming that the independent and dependent variables have a straight-line relationship and independence between variables. This is rarely, if ever, the case in the real world, as there are often many different factors that go into a result (*ML - advantages and disadvantages of linear regression* 2020).

The second model used was Ridge Regression, which solves some of the issues related to the linear regression by "imposing a penalty on the size of the coefficients (*1.1. Linear Models*)." This answers the question of overfitting created by the linear regression model. Ridge is better at identifying the important features and reducing the impact of the less important features, which adds a bias towards the true values that should affect the model. Ridge also has its negatives as well, especially if it is not used in the correct use case. While it can add a bias to make the important inputs more influential, it still uses all the features of the dataset, which can still make the unimportant feature affect the model. While the coefficient theta of the unimportant variable is reduced to a small amount, the amount is still there and present, so it will naturally have an impact on the model, and if they are points that are not needed, it can make the model inaccurate and ultimately unhelpful (Sheth, 2019).

The final model that was looked at was Lasso Regression, which tries to address the downsides of the Ridge model. Lasso does much of the same things that Ridge does in the sense that as values of coefficients increase, a penalty is also placed on that value to decrease the value of the coefficient in order to reduce loss. The big difference between Ridge and Lasso is that while Ridge reduces the value of unimportant features of a dataset, it still gives them a value and are included in the model. Lasso tends to make insignificant values zero, which means they will not affect the model. One of the downsides of this is Lasso can sometimes create a bias for some of the values and become too dependent on those values. Also, if the data set has multicollinear values it will select one of those values randomly, which can misrepresent the data (*Lasso vs Ridge vs elastic net: ML* 2020).

## Data Preprocessing Discussion

For any data project, once a business case is discussed and understood, it is time to find and understand data that is relevant for the business problem. For this specific case, we were given the data in the marketing file. In order to begin understanding the data, it was necessary to do some exploration into the type of data we were dealing with and what its characteristics are. This process began with reading the data into Jupyter Notebook and viewing the first few lines to make sure the data looks as it did on the CSV file that was imported. Then using a describe function we were able to get a better understanding of the count, mean, standard deviation, and other metrics. A couple lines below that we used Sweet Viz to gain more insights into the data, including the range and interquartile range. We also looked at what kind of data was in each column, which was float for all six columns, and check to make sure there were no NULL values.

Once we had a good idea of the data we were working with, it was time to begin preparing the data for the machine learning algorithms we would be using. We started by splitting out the X and Y value, which would be the inputs and outputs for the algorithms. The x

**Commented [MB8]:** https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/

**Commented [MB9]:** https://scikit-learn.org/stable/modules/linear_model.html#ridge-regression

**Commented [MB10]:** https://iq.opengenus.org/ridge-regression/

**Commented [MB11]:** https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/

contains the TV, radio, newspaper, Google, Facebook, and LinkedIn columns, while the y consisted of the sales data. Once that was complete, we split the dataset into test and training datasets, that way our model could be tested on data that was not part of our training set. This allows us fresh data to determine if our model is performing adequately. If we had used data that was in our training set, the result could have been biased due to previous exposure to the data.

Now that the data was separated into test and train sets, we could begin implementing the different models we had selected and see how they performed. We started with the linear regression model then move to the Ridge and finally the Lasso. With each of the algorithms, we first trained the model on the training data set. We then looked at some metrics including at the coefficient of each advertising medium and its effect on sales, then we found the intercept of the model and finally the mean absolute error, mean squared error and root mean squared error for each of the models. In the next sections we will go into those individual metrics and what the result of each of the models when using those metrics.

## Explaining Metrics

Metrics and KPIs are an important part of any business when determining success and successful outcomes. When it comes to regression models that were used in this case, the main metrics that were looked at were the coefficient, the intercept, the mean absolute error, the mean squared error and the root mean squared error. These will help us determine the success of our model and will help us compare the three models to determine which is the best to use for the business case presented.

We start by looking at the size of the coefficient for each of the independent variables, which will show us the effect each independent variable (TV, Radio, newspaper, Google, Facebook, LinkedIn) will have on the dependent variable (Sales). The coefficient number represents the size of the effect, as well as if the effect is negative or positive. This can be read as how much the dependent variable can be expected to increase or decrease depending on the independent variable (*DSS - interpreting regression output*). For this business problem, coefficient is useful to help us understand which advertising mediums have the highest positive effect on sales- this can ultimately help us streamline our allocation of advertising resources based on which medium is most effective. We also look at the intercept of the model to understand where is crosses the x axis and what the value would be if the independent variables were zero.

After the coefficient analysis, we ran our test data through our model to understand how well the model was performing. The items we relied on were the mean absolute error, mean squared error, and root mean squared error. The one we mainly focused on for understanding the model's accuracy was the mean squared error, but we will go through all of them to get a better understanding of each.

**Commented [MB12]:** https://dss.princeton.edu/online_help/analysis/interpreting_regression.htm

We will begin with the mean absolute error, which is the amount of error in the prediction data point and the actual data point. This is found by looking at the absolute residual value, which is the vertical distance of a point to the line for the model, for each point, using the absolute values in order to make sure the values do not cancel out. Once all the residuals have been calculated, we take the mean of those numbers, which will give us the mean distance of the data points to the line. The smaller the mean absolute error the better the model is, with zero being a perfect model. This metric is easily understandable and interpretable, which is its major advantage. The disadvantage is that because we are using the absolute value, it does not indicate if the model is over or under performing and can be affected negatively by outliers (Pascual, 2021).

The second metric that we used to determine the performance of the models was the mean squared error. The mean squared error measures the average squared difference between the predicted value and the actual value. This is used to determine how good the model is at predicting accurate outcomes. Similar to the above, the smaller this number is, the better the model is, and it is useful when comparing different models that are using the same y value; as is the case for this project (Hiregoudar, 2020). Root mean squared error is the square root of mean squared error, which means it is measured in the same units as the target variable, while mean squared error uses units that are the square of the target variable. Root mean squared error serves a similar function as the mean squared error, just using different units, and penalizes larger errors less severely (Dua et al.).

**Commented [MB13]:** https://www.dataquest.io/blog/understanding-regression-error-metrics/

**Commented [MB14]:** https://towardsdatascience.com/ways-to-evaluate-regression-models-77a3ff45ba70

**Commented [MB15]:** https://www.oreilly.com/library/view/machine-learning-with/9781785889936/669125cc-ce5c-4507-a28e-065ebfda8f86.xhtml

## Interpreting Results

Interpreting the results of the models we created will consist of interpreting the results spit out by the model, mainly the coefficient number for each independent variable, and the performance of the model overall as indicated by the mean squared error. We started with the linear regression model, which produced the following coefficients for the different advertising mediums: TV -0.027528, Radio 0.087798, newspaper -0.039369, Google 0.060270, Facebook 0.011841, LinkedIn 0.048429. These coefficient numbers indicate that radio, Google, Facebook, and LinkedIn have a positive impact on sales revenue, and TV and newspaper have a negative impact. It also indicates that radio has the highest positive impact, meaning that radio ads have the highest positive relationship with sales, according to the linear regression model. It is also important to remember the possible downsides of this model, including the impact of outliers. The linear regression had a mean squared error of 4.402118291449691.

The next model that was implemented was the Ridge regression model. Similar to the linear regression model, we used coefficients to determine how effective each medium is. The following are the results from the Ridge model: TV -0.027527, Radio 0.087797, newspaper -0.039368, Google 0.060270, Facebook 0.011841, LinkedIn 0.048429. The interesting thing about the Ridge model coefficients is that they are very similar to the linear model's results, and they had different mean squared errors, with the ridge model being 4.402094649713523.

The mean squared error is better on the Ridge model, which is not super surprising given the Ridge model accounts for some of the cons of the linear model as we looked at earlier. With the coefficient numbers being very similar, the same conclusions can be made: radio and Google look to be the most efficient marketing mediums to create sales, with TV and newspaper still negative.

The final model was the Lasso regression model, and that provided some more answers to the business problem at hand. It produced the following coefficient numbers: TV 0.030607, Radio 0.182945, newspaper -0.001379, Google 0.012935, Facebook 0.001037, LinkedIn -0.000000. These numbers are interesting because they are different than the previous two models in several ways. Starting with TV and radio, which were negative values on the linear and ridge model, the Lasso model still gives newspaper a negative value, but for TV it actually gives it a relatively strong coefficient value. This means it believes TV ads have a positive impact on sales, and the second highest positive value at that. For radio, it lines up with the other models in saying that radio is the most positive impact on sales. Compared to the first two models, it believes Google and Facebook have a positive impact on sales, although the Lasso indicates they do not have as positive of an impact on sales as the Linear and Ridge models indicate. The final independent medium was LinkedIn, which the Linear and Ridge models indicated LinkedIn had a positive impact, but the Lasso set it to zero because it didn't believe it was an important feature. This model does of better job of giving us numbers that we can take a use in our business recommendations. The Lasso model had a mean squared error of 4.393458013886822, which is lower than the previous two models, which indicates that the Lasso model was more accurate that the Ridge and Linear. We will utilize the Lasso model heavily for the business recommendations we make in the next section.

## Recommended Steps

The business problem was to determine which advertising mediums were best at producing sales, recommend the correct mediums, and eliminate those that are not as useful. Looking at all the coefficients from the respective models, it is easy to eliminate newspaper as an effective medium, as all the models indicated it was a negative relationship between newspaper and sales. We can also remove LinkedIn, as the Lasso model deemed it unimportant, and the Ridge and Linear did not have it as a top two positive correlation. Facebook and Google had a positive relationship for all three of the models, which indicates it may be effective, but it is not an overly strong positive relationship. Radio had the strongest positive relationship in all the models, and in the Lasso it came out even stronger than the Ridge and Linear. The one interesting difference between the three was TV, which the Linear and Ridge saw a negative relationship, while the Lasso saw a strong positive. With the discrepancy, we are not confident enough to eliminate TV as a medium. The final recommendation would be the eliminate LinkedIn and newspaper as mediums of advertising, put more resources into radio, as that seems to be the most effective method, and continue to use Google, Facebook, and TV, as there is a positive relationship with sales, and getting more data on these mediums could help us narrow down the results in a future analysis.

# References

*1.1. Linear Models*. scikit. (n.d.). Retrieved January 23, 2022, from https://scikit-learn.org/stable/modules/linear_model.html#ridge-regression

*5 killer examples of data-driven marketing*. WordStream. (2021, December 2). Retrieved January 23, 2022, from https://www.wordstream.com/blog/ws/2016/08/25/data-driven-marketing

Columbus, L. (2018, February 27). *10 Ways Machine Learning is revolutionizing marketing*. Forbes. Retrieved January 23, 2022, from https://www.forbes.com/sites/louiscolumbus/2018/02/25/10-ways-machine-learning-is-revolutionizing-marketing/?sh=3129c0005bb6

*CRISP-DM*. Data Science Process Alliance. (2021, August 24). Retrieved January 23, 2022, from https://www.datascience-pm.com/crisp-dm-2/

Diachuk, O. (2021, July 29). *What is customer lifetime value (CLV)*. owox. Retrieved January 23, 2022, from https://www.owox.com/blog/use-cases/customer-lifetime-value/

Dua, R., Ghotra, M. S., & Pentreath, N. (n.d.). *Machine learning with spark - second edition*. O'Reilly Online Learning. Retrieved January 23, 2022, from https://www.oreilly.com/library/view/machine-learning-with/9781785889936/669125cc-ce5c-4507-a28e-065ebfda8f86.xhtml

Hiregoudar, S. (2020, August 4). *Ways to evaluate regression models*. Medium. Retrieved January 23, 2022, from https://towardsdatascience.com/ways-to-evaluate-regression-models-77a3ff45ba70

*Lasso vs Ridge vs elastic net: ML*. GeeksforGeeks. (2020, March 12). Retrieved January 23, 2022, from https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/

*ML - advantages and disadvantages of linear regression*. GeeksforGeeks. (2020, June 3). Retrieved January 23, 2022, from https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/

Pascual, C. (2021, November 23). *Tutorial: Understanding regression error metrics in python*. Dataquest. Retrieved January 23, 2022, from https://www.dataquest.io/blog/understanding-regression-error-metrics/

Sheth, J. (2019, April 18). *Ridge regression*. OpenGenus IQ: Computing Expertise & Legacy. Retrieved January 23, 2022, from https://iq.opengenus.org/ridge-regression/

*Sklearn.linear_model.linearregression*. scikit. (n.d.). Retrieved January 23, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html?highlight=linear+regression#sklearn.linear_model.LinearRegression

The Trustees of Princeton University. (n.d.). *DSS - interpreting regression output*. Princeton University. Retrieved January 23, 2022, from https://dss.princeton.edu/online_help/analysis/interpreting_regression.htm

Vlada Malysheva, V. O. (2021, May 24). *Machine learning in digital marketing: Examples of use cases: Owox Bi*. owox. Retrieved January 23, 2022, from https://www.owox.com/blog/articles/machine-learning-in-marketing/

*Why is marketing so important*. Moving Targets. (2019, December 4). Retrieved January 23, 2022, from https://movingtargets.com/blog/business-marketing/why-marketing-is-so-important/#:~:text=Marketing%20is%20important%20because%20it%20helps%20you%20sell%20your%20products,is%20ultimately%20what%20drives%20sales.