YEAR	AUTHORS	OBJECTIVES	METHODOLOGY	LIMITATIONS
2014	Kajal kumari	In this article, we will be predicting the prices of used cars. We will be building various Machine Learning models and Deep Learning models with different architectures. In the end, we will see how machine learning models perform in comparison to deep learning models.	In this research, the number of customers is forecasted using machine learning and statistical analysis method with internal data and external data in the ubiquitous environment. Bayesian Linear Regression, Boosted Decision Tree Regression, and Decision Forest Regression are used for machine learning, Stepwise method is used for statistical analysis method. We used Jupyter Notebook as a machine learning tool.	In this article, we tried predicting the car price using the various parameters that were provided in the data about the car. We build machine learning and deep learning models to predict car prices and saw that machine learning-based models performed well at this data than deep learning-based models.

2015	Varun Nayak	Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in the United States	We utilized several classic and state-of-the-art methods, including ensemble learning techniques, with a 90% - 10% split for the training and test data. To reduce the time required for training, we used 500 thousand examples from our dataset. Linear Regression, Random Forest and Gradient Boost were our baseline methods. For most of the model implementations, the open-source Scikit-Learn package [7] was used.	We utilized several classic and state-of-the-art methods, including ensemble learning techniques, with a 90% - 10% split for the training and test data. To reduce the time required for training, we used 500 thousand examples from our dataset. Linear Regression, Random Forest and Gradient Boost were our baseline methods. For most of the model implementations, the open-source Scikit-Learn package [7] was used.

2019	Mukkesh Ganesh	: The production of cars has been steadily increasing in the past decade, with over 70 million passenger cars being produced in the year 2016. This has given rise to the used car market, which on its own has become a booming industry. The recent advent of online portals has facilitated the need for both the customer and the seller to be better informed about the trends and patterns that determine the value of a used car in the market. Using Machine Learning Algorithms such as Lasso Regression, Multiple Regression and Regression trees, we will try to develop a statistical model which will be able to predict the price of a	Overfitting and underfitting come into picture when we create our statistical models. The models might be too biased to the training data and might not perform well on the test data set. This is called overfitting. Likewise, the models might not take into consideration all the variance present in the population and perform poorly on a test data set. This is called underfitting. A perfect balance needs to be achieved between these two, which leads to the concept of Bias-Variance tradeoff. Pierre Geurts [2] has introduced and explained how bias-variance tradeoff is achieved in both regression and classification. The selection of variables/attribute plays a vital role in influencing both the bias and variance of the statistical model. Robert Tibshirani [3] proposed a new method called Lasso, which minimizes the residual sum of squares. This returns a subset of attributes which need to be included in multiple regression to get the minimal error rate. Similarly, decision trees suffer from overfitting if they are not pruned/shrunk. Trevor Hastie and Daryl Pregibon [4] have explained the concept of pruning in their research paper. Moreover, hypothesis testing using ANOVA is needed to verify whether the	The prediction error rate of all the models was well under the accepted 5% of error. But, on further analysis, the mean error of the regression tree model was found to be more than the mean error rate of the multiple regression and lasso regression models. Even though for some seeds the regression tree has better accuracy, its error rates are higher for the rest. This has been confirmed by performing an ANOVA. Also, the post-hoc test revealed that the error rates in multiple regression models and lasso regression models aren’t significantly different from each other. To get even more accurate models, we can also choose more advanced machine learning algorithms such as random forests, an ensemble learning algorithm which creates multiple decision/regression trees, which brings down overfitting massively or Boosting, which

		used car, based on previous consumer data and a given set of features. We will also be comparing the prediction accuracy of these models to determine the optimal one.	different groups of errors really differ from each other. This is explained by TK Kim and Tae Kyun in their paper [5]. A Post-Hoc test needs to be performed along with ANOVA if the number of groups exceeds two.	tries to bias the overall model by weighing in the favor of good performers. More data from newer websites and different countries can also be scraped and this data can be used to retrain these models to check for reproducibility.

		specific growth of each SKU (Stock Keeping Unit), companies can fall into the traps of subjectivity or generalism.	accuracy and interpretability of the statistical model it produces. Lasso regression could also be a kind of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, a bit like the mean. The lasso procedure encourages simple, sparse models. This particular kind of regression is well-suited once we would like to automate certain parts of model selection, like variable selection/parameter elimination.	customer behaviour. For fresh produce, forecast accuracy is highest when done for the shortest time intervals as this diminishes in accuracy with long term future predictions.

		geographic / regional preferences and changes in consumer behavior. We understand that the need for a predictive approach through data, mainly external to the company, is increasingly latent.	assistive calorie measurement system. In proposed a time-dependent food distribution model and a weight optimisation algorithm aimed at adapting the user's data to their eating habits. Deep learning has also been imposed in the waste sorting process to automate some of the waste handling tasks.	to think that this is linked to demand shifts and believe it is the true cause of it.