Prediction of Workforce Required for a Laundry Firm Using Machine Learning Algorithms

Abhilash, Anushree Anand, Hitayshi P Makadia, Mahima Krishna

Assistant Professor, Student, Student, Student
Information Science and Engineering,
Atria Institute of Technology, Bangalore, India.

Abstract: For most people, washing laundry at home is a good option. But there are times where people may not be able to do it because of several reasons like work schedules that leave you too exhausted to even do your own laundry. While doing your own laundry might be a good option for saving some money, it can be time consuming. Choosing a laundry service for your clothes and other items has several advantages. Laundry establishments face difficulty in maintaining detailed records of customer clothing; this little problem as seen in most laundry firms discourages customers. In the current system used by laundry firms, several problems arise: customer clothes mix-up, untimely pick-up, and drop-off of clothes, billing related issues, and so on. The aim of this application is to digitalize the current system which will make laundry firms more efficient and ensure customer satisfaction. In this paper, machine learning algorithms are used to build a prediction model which can be used to determine the workforce required to help the firm allocate and utilize resources effectively and efficiently. This study compares two different machine learning algorithms: Random Forest and Bayesian Regression from Linear Models to determine which algorithm predicts the output more accurately. A database was created after conducting a detailed survey and interacting with various firms to understand the inflow and outflow of clothing items and other garments and how they were handled: right from collection to delivery. The resulting database was used to train and test the above-mentioned algorithms. This paper provides information about which machine learning algorithm may be incorporated with existing commercial laundry applications as an additional feature to improve their respective businesses.

***IndexTerms* - Laundry Management, Machine learning, Prediction, Random Forest, Bayesian regression, Linear models.**

I. INTRODUCTION

There are several applications of machine learning, the foremost significant of which is data processing. People are often at risk of making mistakes during analyses or, possibly, when trying to determine relationships between multiple features. This makes it difficult for them to seek out solutions to certain problems. Machine learning can often be successfully applied to those problems, improving efficiency of systems and therefore the design of machines. The two machine learning algorithms being used here are based on supervised learning techniques, they are, the random forest algorithm and bayesian regression. The goal of supervised learning is to create a concise model of the distribution of sophistication labels in terms of predictor features. The resulting classifier is then wont to assign class labels to the testing instances where the values of the predictor features are known, but the value of the category label is unknown.

Traditionally, researchers have used either off-the-shelf models like COCOMO, or developed local models using statistical techniques like stepwise regression, to get software effort estimates. More recently, attention has turned to a spread of machine learning methods like artificial neural networks (ANNs), case-based reasoning (CBR) and rule induction (RI).

Every business seeks to grow. But just some of the businesses that successfully actualize this vision do so through a data-based higher knowledge. And to make these informed decisions, companies are using machine learning-based predictive analytics.

Predictive analytics is predicting future outcomes supported by historical and current data. It uses various statistical and data modelling techniques to research past data, identify trends, and help make informed business decisions. While previously, machine learning and predictive analytics were viewed as two entirely different and unrelated concepts, the increasing demands of effective data analytics have brought machine learning algorithms to intertwine with predictive analytics. Today, predictive analytics extensively uses machine learning for data modelling thanks to its ability to accurately process vast amounts of data and recognize patterns.

When it's unattainable to compute the complete result earlier, due to uncertain or unforeseeable events, model predictive control methods are often applied to manage the system during operation in real-time.

Predictive modelling, also called predictive analysis, may be a computing that seeks to predict future events or outcomes by analyzing patterns that are likely to forecast future results. The goal of predictive modelling is to answer this question: “Based on known past behaviour, what's presumably to happen within the future?”.

Increasing awareness of the importance of environmental protection has strongly influenced the economic landscape during the last decade. As a result, there's a general desire to scale back the consumption of energy and resources. This can be also a central issue within the development of commercial laundries, where, particularly, reduction of energy consumption and also the number of washing detergents are of interest. Positive side effects include the reduction of operating costs. Ideally, one wants to maximize economic efficiency while minimizing the consumption of energy and resources. In this article, we present a prediction model for laundry and cleansing establishments to assist improve their efficiency and optimize the use of resources.

The database used here consists of 500 rows of data, where each row represents data collected from a group of establishments. The database shows the number of items of clothing collected, and out of these collected items for various services like dry cleaning, ironing, darning, etc. Out of the clothes collected, it shows the number of garments on which services have been performed and then delivered back to the customer successfully. There is one column which represents the number of people(or the workforce) required to finish working on the specified number of clothing items.

Two algorithms to calculate the workforce required namely and then determine which algorithm gives a more accurate prediction of the workforce required.

II. General Structure

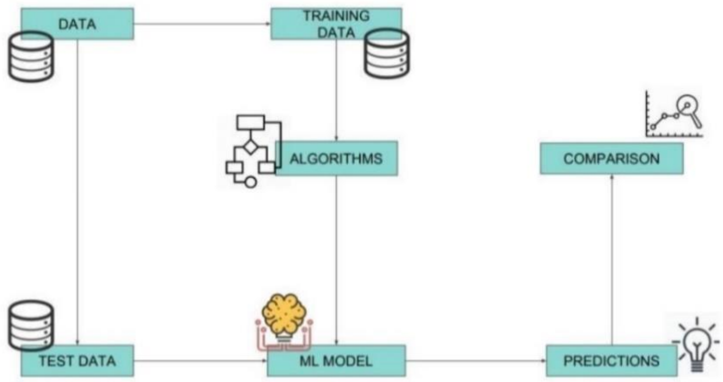

Fig. 1 General Structure of Machine Learning Model

The above figure represents the general structure of an ML model. The data are split into Training and Test Data, respectively. The Training Data is passed to trough the ML algorithms for enabling the machine to learn and apply it on the test data to predict the solutions. The such predicted solutions can be used for comparisons, calculate the accuracy of the prediction.

III. Related Survey

3.1 Past research on Prediction using Random Forest approach

The Random Forrest approach included the use of decision tree pruning, a kind of typical single classifier, the first step of training set is to recursive analysis, generate a shape such as inverted tree structure; The second step analysis of the tree from the root node to leaf node path, produce a series of rules. Finally, according to these rules, classification, or projections for new data. Every tree in the forest depend on a random vector, vector in the forest are all independent identically distributed.

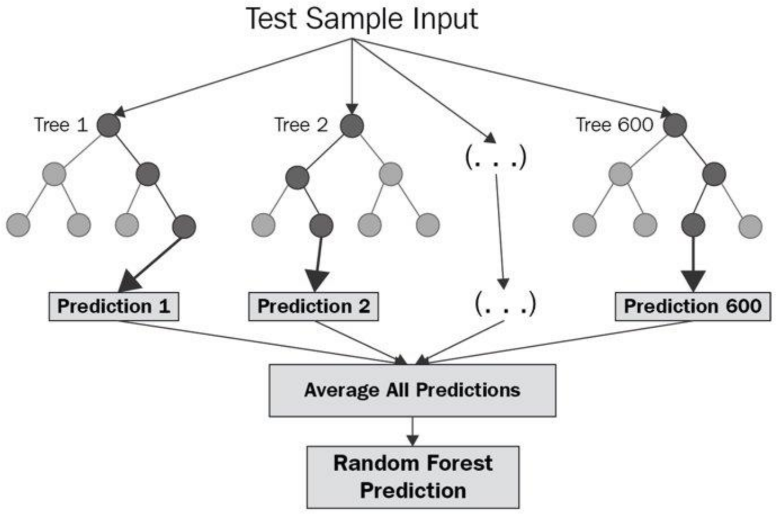

Fig. 2 Random Forest Structure

Random Forest is a meta-algorithm that combines a large number of decision-tree models, each individually built on bootstrapped samples of the data. This process of sampling the data and combining the individual decision-trees is called Bagging and is able to reduce the variance of the predictions without increasing the bias. The predictions are formed by taking the mean of the individual decision-tree predictions. The random forest method along with enhancement with Tuner is used.

From a computational standpoint, Random Forests are appealing because they

- naturally handle both regression and classification.
- are relatively fast to train and to predict.
- depend only on one or two tuning parameters.
- have a built-in estimate of generalization error.
- can be used directly for high-dimensional problems.
- can easily be implemented in parallel.

Statistically, Random Forests are appealing because of the additional features they provide, such as,

- measures of variable importance.
- differential class weighting.
- missing value imputation.
- visualization.

3.2 Past research on Prediction, using Bayesian Linear Regression Approach

In statistics, Bayesian regression is an approach to regression within which we undertake the applied mathematics analysis at intervals in Bayesian reasoning. Once the regression model has errors that have a traditional distribution, and we assume if a selected variety of previous distribution, specific results are obtainable for the posterior likelihood distributions of the model's parameters.

Linear models describe a continual response variable as a performance of one or additional predictor variables. They assist you perceive and predict the behavior of advanced systems or analyze experimental, financial, and biological knowledge. Regression may be a statistical procedure won't to produce a linear model. The model describes the link between a variable quantity y (also referred to as the response) as a performance of 1 or additional freelance variables Xi (called the predictors). the final equation for a linear model is: $y=\beta 0+\sum \beta iXi+\epsilon i$ wherever represents the linear parameter estimates to be computed and ϵ represents the error terms. In the Bayesian viewpoint, we have a tendency to plan regression victimisation likelihood distributions instead of purpose estimates. The response, y, isn't calculable as one price, however, is assumed to be drawn from a likelihood of distribution. The model for Bayesian regression with the response sampled from a traditional distribution is:

$$y \sim N(\beta^T X, \sigma^2 I)$$

It generates the output y from a traditional (Gaussian) Distribution characterized by a mean and variance. The mean for regressionis that the transpose of the load matrix increased by the predictor matrix. The variance is that the sq. of the quality the quality (multiplied by the scalar matrix because of this is often a multi-dimensional formulation of the model).

The aim of Bayesian regression isn't to seek the one "best" price of the model parameters, however, rather to see the posterior distribution for the model parameters. Not solely is that the response generated from a likelihood of distribution, however the model parameters are assumed to come back from a distribution also. The posterior likelihood of the model parameters is conditional upon the coaching inputs and outputs:

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

Here, P(β|y, X) is that the posterior likelihood distribution of the model parameters given the inputs and outputs. This is often adequate to the chance of the information, P(y|β, X), increased by the previous likelihood of the parameters and divided by a normalization constant. This is often a simple expression of Bayes Theorem, the elemental underpinning of Bayesian Inference:

$$Posterior = \frac{Likelihood * Prior}{Normalization}$$

Let's stop and place confidence in what this implies. In distinction to OLS, we've a posterior distribution for the model parameters that's proportional to the chance of the information increased by the previous likelihood of the parameters. Here we can observe the two primary advantages of Bayesian regression.

1. Priors: If we've domain data, or a guess for what the model parameters ought to be, we can embrace them in our model, in contrast to within the frequentist approach that assumes everything there's to grasp regarding the parameters comes from the information. If we have a tendency to don't have any previous estimates, we can use non-informative priors for the parameters like a traditional distribution.
2. Posterior: The results of acting Bayesian regression may be a distribution of potential model parameters supported by the information and also the previous. this permits the quantity our uncertainty regarding the model: if we've less knowledge points, the posterior distribution are additional unfolded.

As the quantity of information points will increase, the chance washes out the previous, and within the case of infinite knowledge, the outputs for the parameters converge to the values got from OLS.

IV. RESULTS AND DISCUSSION

Results by using Random Forest Algorithm

In the figure 3, the graph of actual values versus predicted values for random forest is shown. It can be observed that there is a large deviation between the actual and predicted values and therefore, this algorithm is not very accurate.

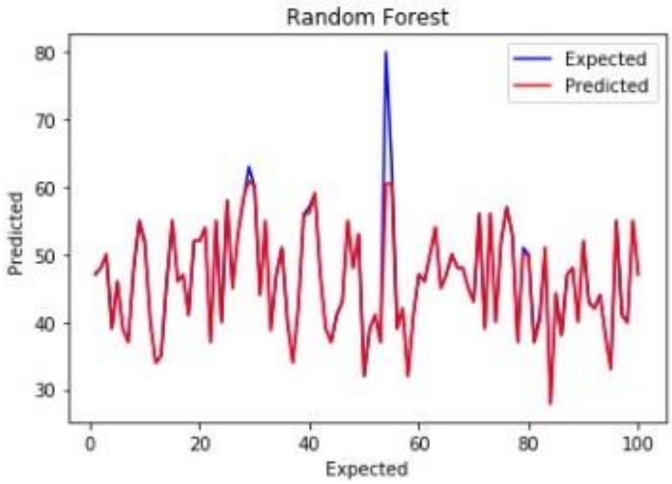

Fig. 3 Graph using Random Forest Prediction Algorithm

Results by using Linear Bayesian Ridge

Figure 4 represents the graph of actual values versus predicted values for Bayesian regression. It can be observed that there is very little deviation between the actual and predicted values and therefore, this algorithm is very accurate.

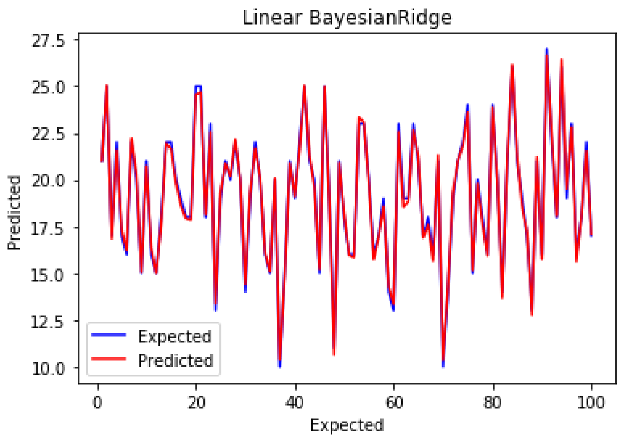

Fig. 4 Graph using Linear Bayesian Ridge Prediction Algorithm

Accuracy comparison of the two algorithms

Model	Accuracy
Random Forest Algorithm	94.1%
Linear Bayesian Ridge	98.2%

Table. 1 Accuracy comparisons of the algorithms

V.CONCLUSION

This paper provides information about which machine learning algorithm may be incorporated with existing commercial laundry applications as an additional feature to improve their respective businesses. From the above comparisons, it is clear that the Bayesian regression model is more accurate than the random forest model. Increasing awareness of the importance of environmental protection has strongly influenced the economic landscape during the last decade. As a result, there's a general desire to scale back the consumption of energy and resources. This can be also a central issue within the development of commercial laundries, where, particularly, reduction of energy consumption and also the number of washing detergents are of interest. Our approach to the laundry management process helps laundry establishments tackle just this. By understanding the expected number of clothing and predicting the workforce required, resources can be allocated utilized effectively and efficiently.

REFERENCES

[1] Bayesian linear regression Thomas P. Minka 1998 (revised 2010)
[2] Random Forests Leo Breiman Statistics Department University of California Berkeley, CA 94720 January 2001
[3] Bayesian Regression: https://scikit-learn.org/stable/modules/linear_model.html#bayesian-regression
[4] RandomForestRegressor: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html