


# Supervised Machine Learning for Predicting SMME Sales: An Evaluation of Three Algorithms

## Helper Zhou

PhD Candidate, Department of Entrepreneurial Studies and Management, Durban University of Technology

 <https://orcid.org/0000-0002-8492-7844>

## Victor Gumbo

Senior Lecturer, Department of Mathematics, University of Botswana

 <https://orcid.org/0000-0001-5219-9902>

## Abstract

The emergence of machine learning algorithms presents the opportunity for a variety of stakeholders to perform advanced predictive analytics and to make informed decisions. However, to date there have been few studies in developing countries that evaluate the performance of such algorithms—with the result that pertinent stakeholders lack an informed basis for selecting appropriate techniques for modelling tasks. This study aims to address this gap by evaluating the performance of three machine learning techniques: ordinary least squares (OLS), least absolute shrinkage and selection operator (LASSO), and artificial neural networks (ANNs). These techniques are evaluated in respect of their ability to perform predictive modelling of the sales performance of small, medium and micro enterprises (SMMEs) engaged in manufacturing. The evaluation finds that the ANNs algorithm's performance is far superior to that of the other two techniques, OLS and LASSO, in predicting the SMMEs' sales performance.

## Keywords

supervised machine learning, algorithms, sales predictive modelling, ordinary least squares (OLS), least absolute shrinkage and selection operator (LASSO), artificial neural networks (ANNs), small, medium and micro enterprises (SMMEs)

**DOI:** <https://doi.org/10.23962/10539/31371>

## Recommended citation

Zhou, H., & Gumbo, V. (2021). Supervised machine learning for predicting SMME sales: An evaluation of three algorithms. *The African Journal of Information and Communication (AJIC)*, 27, 1-21. <https://doi.org/10.23962/10539/31371>



This article is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence: <https://creativecommons.org/licenses/by/4.0>

## 1. Introduction

Today's organisations, both small and large, handle increasingly large amounts of data, and the amounts are expected to continue to grow exponentially (Cheriyana et al., 2018; Ndikum, 2020). Ndikum (2020) notes that human beings generate and store in excess of 2.5 quintillion bytes of data daily. Inevitably, the availability of such huge amounts of data has provided an impetus for organisations to harness efficient and flexible methods to conduct predictive analytics and inform data-driven future plans (Bajari et al., 2015; Leo et al., 2019; Obaid et al., 2018).

Machine learning techniques are attracting the interest of numerous stakeholders, including private-sector entities seeking the means to intelligently exploit their data to aid decision-making and enhance their competitive advantage in the market (Dod & Sharma, 2010; Krishna et al., 2017; Tsoumakas, 2019). Kolkman and Van Witteloostuijn (2019) explain that machine learning enables businesses to perform advanced predictive modelling to an extent not possible with traditional statistical techniques (Leo et al., 2019; Van Liebergen, 2017). Machine learning has been widely embraced for a variety of purposes, including financial modelling, health and safety analysis, medical diagnosis, and fraud detection (Crane-Droesch, 2017; Enkono & Suresh, 2020; Gholizadeh et al., 2018; Mohammed et al., 2016). Machine learning techniques have also been embraced for predicting market demand and consumer behaviour (Bajari et al., 2015; Sekban, 2019; Tsoumakas, 2019; Venishetty, 2019). The power of machine learning has attracted significant interest from numerous players, including business owners, data scientists, and econometricians (Bajari et al., 2015; Sekban, 2019; Venishetty, 2019).

Sales predictions are one of the most important elements of business operations, including for small-sized firms seeking to sustainably increase sales in order to enhance their chances of survival (Sekban, 2019; Venishetty, 2019). The rise of advanced data analytics techniques provides SMMEs with opportunities to conduct sales performance predictive modelling (Krishna et al., 2017; Tsoumakas, 2019). However, despite their significant contribution to predictive analytics, machine learning techniques have not yet been fully exploited in small enterprises' research and practice. The existing literature provides very few studies on SMMEs' use of machine learning in developed countries or in developing countries such as South Africa (Bauer, 2020; Haataja, 2016; Kolkman & Van Witteloostuijn, 2019; Te, 2018).

In respect of machine learning algorithms, Ryll and Seidens (2019) note that the extant literature lacks an evaluation of the various algorithms' effectiveness. The result is that stakeholders are likely to arbitrarily select an algorithm, without any scientific basis for their choice. Identification of the best-performing predictive techniques for particular settings and purposes would provide stakeholders with bases for deciding which to use.

To address this gap in the South African context, our study evaluated the performance of three supervised machine learning algorithms that can be used to conduct sales predictive modelling: OLS, LASSO, and ANNs. The algorithms' ability to predict SMME sales performance was evaluated using a panel dataset of manufacturing SMMEs in South Africa's KwaZulu-Natal (KZN) Province.

## 2. Machine learning

According to Ryll and Seidens (2019), the concept of machine learning, despite its growing popularity, remains ill-defined in extant literature. The authors define it as a process through which a system interacts with its environment in such a way that the system's structure changes and, owing to structural alterations, the interaction process changes as well. Shalev-Shwartz and Ben-David (2014) assert that machine learning is the detection of meaningful data patterns by algorithms in an automated way, essentially indicating that machine learning techniques endow programs with the ability to "learn" and adjust accordingly. This conception aligns with that of Goodfellow et al. (2016), who define machine learning as the ability of artificial intelligence (AI) systems to acquire knowledge by gleaning patterns from raw datasets. Lantz (2019) conceives machine learning as being concerned with techniques that process and transform data into actionable intelligence. Mohammed et al. (2016) describes machine learning as the enablement of machines to learn without explicit programming.

A key advantage of various machine learning techniques like ANNs is that they are non-parametric, i.e., they do not require features in the dataset to be normally distributed, as do some classical statistical modelling approaches (Kolkman & Van Witteloostuijn, 2019; Van Liebergen, 2017). This flexibility allows algorithms to learn, adapt, and in the process uncover subtle insights in data (Leo et al., 2019).

Research has shown that organisations which adopt machine learning algorithms for predictive modelling will benefit in many ways, including more effective strategic planning, resource optimisation, risk management, and inevitably enhanced competitive advantage (Cheriyana et al., 2018; Kolkman & Van Witteloostuijn, 2019; Leo et al., 2019). Krishna et al. (2017) have found that algorithms can be used to accelerate business performance and achieve long-term goals. One of the main areas in which machine learning techniques have been used is in sales performance predictive modelling (Sekban, 2019; Tsoumakas, 2019; Venishetty, 2019). This is because sales directly impact enterprise survival and long-term growth (Bauer, 2020; Sekban, 2019).

### *Supervised machine learning*

Machine learning techniques can be of either a supervised or unsupervised nature. *Unsupervised* techniques are used when dealing with unlabelled datasets (Mohammed et al., 2016; Venishetty, 2019). In unsupervised learning, the interest is more in the structure of the dataset as it is analysed, without specifying a response variable to predict (Aziz & Dowling, 2019; Mohammed et al., 2016; Van Liebergen, 2017).

*Supervised* techniques are used when features in the dataset are labelled and the target variable is known and specified (Ryll & Seidens, 2019; Venishetty, 2019). In this study, the techniques used fall under the supervised paradigm.

Under the supervised machine learning paradigm, tasks are grouped into either classification or regression (Venishetty, 2019). *Classification* can be used, for instance, to predict (in this case) whether an SMME will grow (1) or not grow (0) in the next year, and this type of task is commonly termed a binary classification. On the other hand, *regression* tasks involve the prediction of a continuous variable, like (in this case) the prediction of an SMME's sales.

To ensure enhanced model performance, the common practice is to conduct *data partitioning*, i.e., dividing the data into two separate parts, commonly known as the *training* and *test* datasets (Bauer, 2020). *Training* data, which is labelled and thus “seen”, is used for model-building, and the *test* data, which is unlabelled and thus “unseen”, is used for model validation or testing (Mohammed et al., 2016; Te, 2018). This partitioning allows algorithms fitting well on training data to be checked to make sure they are not “overfitting” when applied to the test data (Mohammed et al., 2016). (Some algorithms might fare well on the training (seen) data but poorly on the test (unseen) data, and this is known as overfitting.) The training dataset is made up of input vector  $X$  and output vector  $Y$ , both of which have labelled features. In the training phase, algorithms learn to approximate a function to produce  $\hat{Y}$  which is also denoted  $\hat{f}(X)$ . Thus, through using different algorithms, as per Equation (1) below, a mapping function from  $X$  to  $Y$  is learned.

$$Y = f(X) + \varepsilon \quad (1)$$

Based on Equation (1),  $\varepsilon$  is the error term independent of the explanatory variables, and despite the performance of the mapping function this error cannot be reduced.

### ***Supervised machine learning tools***

Choosing an appropriate algorithm for any given task is not a trifling decision but an important one, because the results from the selected technique will influence and guide decision-making. As argued by Venishetty (2019), there is no “one-size-fits-all” machine learning technique for every problem and thus there is a need to evaluate and identify an appropriate algorithm for a given task. Various machine learning techniques have been used to solve regression problems such as sales modelling. OLS, LASSO, and ANNs are among the most extensively used algorithms for such learning tasks (Casella et al., 2017; Lantz, 2019; Melkumova & Shatskikh, 2017; Shalev-Shwartz & Ben-David, 2014).

*Ordinary least squares (OLS)*

The OLS technique, which is also generally referred to as the linear regression technique, is valued mainly for its ability to learn efficiently. It has been found to provide linear predictors that are not only intuitive and easily interpretable, but also perform reasonably well in fitting data in different natural learning problems (Casella et al., 2017; Shalev-Shwartz & Ben-David, 2014). This form of predictive technique is normally used in traditional statistical modelling when ascertaining causal relationships between response variables and dependent variables (Aziz & Dowling, 2019). In essence, this technique attempts to choose the slope and the intercept that minimise the sum of the squared errors—or, as described by Lantz (2019), to minimise the distance between the predicted and the actual target variable.

Expressed in mathematical terms, the goal of OLS regression modelling is to minimise the error ( $e$ ), also known as the sum of squared residuals, which is the difference between predicted value  $\hat{Y}$  and the actual value  $y$  as per Equation (2):

$$\sum (y_i - \hat{y}_i)^2 = \sum e_i^2 \quad (2)$$

As can be noted in Equation (2), in order to eliminate negative values, the error values are squared and summed across all data points.

Key shortcomings with OLS are its linearity assumption between the response and predictor variables and its inability to deal with collinearity (Kolkman & Van Witteloostuijn, 2019; Van Liebergen, 2017). Nonetheless, OLS is one of the most popular techniques in academic research. Kolkman and Van Witteloostuijn (2019) describe OLS as the empirical “workhorse” in academia. The algorithm was included in this study as the traditional benchmark so as to enable cross-method comparisons with the two other algorithms evaluated.

*Least absolute shrinkage and selection operator (LASSO)*

The LASSO method is mainly used to achieve simultaneous parameter estimation and model selection in regression analysis (Muthukrishnan & Rohini, 2016). This algorithm zero weights covariates with low explanatory power and allows one to work with an interpretable parsimonious model (Aziz & Dowling, 2019; Leo et al., 2019; Melkumova & Shatskikh, 2017). Casella et al. (2017) find that the LASSO technique performs better than OLS, and another related technique called ridge regression, in predictive analytics. The LASSO technique shares similarities with OLS, save that, unlike the latter, LASSO employs the  $\ell_1$  penalty function. In essence,

LASSO is a simple OLS technique with feature selection and regularisation embedded in it. Following Muthukrishnan and Rohini (2016), we defined our LASSO estimates as per Equation (3) below:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (3)$$

Based on Equation (3),  $\lambda \geq 0$  is a tuning parameter, and when  $\lambda = 0$  the penalty has no effect and LASSO will produce similar estimates to those of least squares. However,  $\lambda \rightarrow \infty$ ,  $\ell_1$  forces some of the coefficient estimates to zero, thereby performing forward-looking variable selection. LASSO effectively deals with the problem of collinearity among predictors by selecting only one and shrinking other variables to zero, thereby producing stable and accurate predictions (Casella et al., 2017; Muthukrishnan & Rohini, 2016).

#### *Artificial neural networks (ANNs)*

ANN algorithms are inspired by the structure of the internal functioning of the human brain and nervous system (Shalev-Shwartz & Ben-David, 2014). The technique aims to solve problems by mimicking the human brain, through learning from past experiences and then making use of those learnings as a basis for making future decisions. This technique differs from traditional statistical techniques in that it is non-parametric, i.e., it makes no presumptions on the data distribution (Youn & Gu, 2010). ANN algorithms have become popular for implementing machine learning (Krishna et al., 2017) owing to their ability to yield an effective learning paradigm that produces excellent performance on various learning tasks (Shalev-Shwartz & Ben-David, 2014). The neural network is a network of connected nodes, and for each node, inputs are summed before being linearly transformed. Equation (4) presents an ANN mathematically:

$$o = \sum_{i=1}^n w_i x_i + b \quad (4)$$

Where  $x_i$  is the  $i$ th input to the ANN node,  $w_i$  the  $i$ th input weight,  $n$  the number of inputs,  $b$  the bias term and  $o$  the node output.

The motivations for the adoption of this technique include its flexibility—in increasingly complex data structures—in addressing outliers, missing data, multicollinearity, and nonlinearities (Gepp & Kumar, 2012; Merkel et al., 2018). The advantage of ANN algorithms lies in their versatility, as they can be applied to virtually any learning task,



be it regression, classification, or even unsupervised learning tasks (Leo et al., 2019; Youn & Gu, 2010). The class of ANN we used in this study is the multilayer perceptron (MLP), which is also referred to as a multilayer feedforward network (Lantz, 2019).

### *Existing comparative findings on the three tools*

Findings reported in the existing literature shown that, generally, in terms of predictive performance across different fields, ANN algorithms perform better than OLS. Nghiep and Al (2001) find that compared to the OLS technique, ANNs performed better in predicting residential property value. This finding is in line with the Farahani et al. (2016) study, which evaluates the performance of ANNs and OLS techniques to predict car sales and finds ANNs superior. Ahangar et al. (2010) establish the superior performance of ANNs compared to OLS in predicting the stock price of listed companies. Croda et al. (2019) establish that ANNs have a very high predictive accuracy compared to traditional statistical techniques in sales forecasting, even when presented with a small dataset. Accordingly, alternative methods aiming to improve OLS, such as the LASSO technique as per Equation (3), have been established (Casella et al., 2017; Tibshirani, 2011).

Ratnasena et al. (2021) find that compared to ANNs, the LASSO technique more accurately predicts the condition of tapes in sampled US cultural heritage institutions. Das et al. (2018) find that LASSO performs better than both ANNs and (as expected) OLS in predicting rice yields in India. Castelli et al. (2020) find that LASSO is more accurate than ANNs in predicting online property trends in Bulgaria. Utilising European Environmental Agency air pollution data, Chen et al. (2019) find that both LASSO and OLS have superior predictive performance compared to the ANNs in predicting the annual average concentration of fine particle and nitrogen dioxide across Europe.

Strandberg and Låås (2019), using data on Swedish companies, find that ANNs perform significantly better than LASSO in predicting sales performance.

Droomer and Bekker (2020), utilising a large database of US online grocery stores, find that ANNs outperform other modern and complex algorithms like XGBoost in predicting customers' purchasing behaviour. Croda et al. (2019), using a small Mexican chemicals wholesaler dataset, establish that ANNs produce highly accurate sales predictions. Wang et al. (2019) demonstrate the high accuracy of ANNs in predicting the annual sales of Taiwanese manufacturing enterprises. Penpece and Elma (2014) show that ANNs produce sales predictions that are close to the actual data of Turkish retail stores.

### 3. Study design and methodology

The study used R version 3.6.3, an open source software developed by the R Development Core Team (2019).

#### *Dataset preparation*

The three-year longitudinal dataset, containing information on 191 manufacturing SMMEs in KwaZulu-Natal Province for the years 2015 to 2017, was accessed from McFah Consultancy, a Durban-based company focusing on business and tax advisory services for SMMEs. The majority of the SMMEs (61%) in the dataset were from eThekweni Metropolitan Municipality (greater Durban), followed by King Cetshwayo District (11%), uThukela District (10%), uMgungundlovu District (7%), iLembe District (3%), Amajuba District (3%), Ugu District (2%), Zululand District (2%), uMzinyathi District (1%) and uMkhanyakude District (1%). There were no SMMEs from Harry Gwala District. The data had the following features: sales, owner's gender, enterprise location, owner's year of birth, total assets value, permanent employees, temporary employees, digital marketing medium use, website use, enterprise registration type, and registration year. Three macroeconomic variables were also included in the dataset: gross domestic product (GDP) and unemployment statistics from Statistics South Africa (2018), and the purchasing managers' index (PMI) from the Bureau for Economic Research (n.d.).

#### *Target variable*

Since the interest, for this study, was in evaluating the predictive potency of OLS, LASSO, and ANNs with respect to enterprise performance, it was important to define the target variable based on the dataset. In line with previous studies, enterprise performance was proxied by sales (Buyinza, 2011; Panda, 2015; Phillipson et al., 2019), which we coded as LogSales.

#### *Independent variables*

Total assets were coded as LogTA, total number of permanent workers as Pemp, number of temporary workers as Temp, and labour productivity, which was measured by sales per employee, was coded as Prod. The SMME owner's gender—proxied by 1 for male and 0 for female—was coded as Gen. The SMME owner's age, which was measured as the difference between the panel dataset period (2015 to 2017) and year of birth, was coded as EntAge. Having a website—proxied by 1 for enterprises with a website and 0 for those without—was coded as Web. The company's age, which was coded as CoAge, was measured as the difference between the panel data period and



the year of registration. The SMME's registration type, which was the legal structure of the participating enterprises, was defined by 1 for limited liability registered enterprises and 0 for other, and this variable was coded as Reg. For digital marketing, coded as DigMkt, the dummy variable 1 was used for those using one or more of three digital marketing platforms (Facebook, Twitter, and Instagram) and 0 for those not using any of these platforms. SMME location, coded as Loc, was proxied by 1 for those based in eThekweni Metropolitan Municipality and 0 for those located in district municipalities.

Three additional polynomial features were constructed to assess the nonlinear effects of these variables on enterprise performance. These were the owner's age squared (EntAge2), the SMME's age squared (CoAge2), and temporary workers squared (Temp2).

Finally, three external variables were used: the national annual economic growth rate, coded as GDP; the national unemployment rate, coded as Unemp; and the purchasing managers' index, coded as PMI, calculated as the average annual PMI rate for each of the three years between 2015 and 2017. Exploratory analysis showed that the dataset was not stationary; to address this, we followed Curran-Everett (2018) by log transforming all continuous variables (i.e., sales, total assets, permanent workers, temporary workers, productivity, owners' ages, and SMMEs' ages). Consequently, the transformation stabilised the variance of all continuous variables.

### *Hypothesis-testing*

Model-building was done after conducting hypotheses tests to establish variables with an impact on sales performance. Hypothesis testing is an important step in model building, as this enables the identification of key factors which impact the target variable (Punam et al., 2018). The benefit of this step is that the data features selected for training the algorithm are those that best explains sales performance, and irrelevant features, which tend to adversely impact model accuracy due to data redundancy, are removed. Furthermore, a model built using important variables tends to minimise the challenge of overfitting, the model training time is significantly reduced, and overall, the model performs better when applied to real world problems (Venishetty, 2019). An in-depth literature review was thus conducted, and Table 1 provides the 13 hypotheses that were derived for empirical investigation to identify features with a significant effect on SMMEs' sales performance that were then used for model-building.

**Table 1: Hypotheses**

Hypothesis	Supporting literature
H1: Entrepreneur's gender has a significant effect on enterprise performance	(Amran, 2011; Bardasi et al., 2011; Essel et al., 2019)
H2: Entrepreneur's age has a significant nonlinear effect on enterprise performance	(Amran, 2011; De Kok et al., 2010; Kaunda, 2013)
H3: Labour productivity has a significant effect on enterprise performance	(Bellone et al., 2008; Bigsten & Gebreyesus, 2007; Esteve-Pérez & Mañez-Castillejo, 2006)
H4: Permanent employees have a significant effect on enterprise performance	(Clinebell & Clinebell, 2007; Pauka, 2015; Thorsteinson, 2003)
H5: Temporary employees have a significant nonlinear effect on enterprise performance	(Chadwick & Flinchbaugh, 2016; Pauka, 2015; Roca-Puig et al., 2012;)
H6: Total assets have a significant effect on enterprise performance	(Al-Ani, 2013; Gupta et al., 2013; Maggina & Tsaklanganos, 2012)
H7: SMME's age has a significant nonlinear effect on enterprise performance	(Coad et al., 2018; Loderer & Waelchli, 2010; Rijkers et al., 2010)
H8: Limited liability registration type has a significant effect on enterprise performance	(Adegbite et al., 2007; Muriithi, 2017; Small Business Project, 2014;)
H9: Usage of digital media platforms significantly impacts enterprise performance	(Camilleri, 2018; Jobs & Gilfoil, 2014; Parsons, 2013)
H10: Website usage has a significant effect on enterprise performance	(Jobs & Gilfoil, 2014; Meroño-Cerdan & Soto-Acosta, 2005; Parsons, 2013)
H11: National unemployment rate has a significant effect on enterprise performance	(Halicioglu & Yolac, 2015; Huggins et al., 2017)
H12: GDP growth rate has a significant effect on enterprise performance	(Egbunike & Okerekeoti, 2018; Klapper & Richmond, 2011; Motoki & Gutierrez, 2015)
H13: PMI has a significant effect on enterprise performance	(Harris, 1991; Koenig, 2002)

Note: Based on the consulted literature, three of the variables—entrepreneur's age, temporary workers and SMME's age—were each expected to have a nonlinear effect on performance. This suggests that over time, unlike the other 10 variables, each of these three was expected to have a turning point, i.e., a point when the variable switches from having a negative effect to having a positive effect (or vice versa) on firm performance.

To empirically test the above hypotheses, the random effects within between (REWB) panel data modelling technique (Bell et al., 2019) was used. The distinct

advantage of REWB over other techniques, such as fixed effects or random effects, is that the former simultaneously captures both micro and macro associations of the independent variables on the target variable (Bell & Jones, 2015; Bell et al., 2019).

The hypotheses-testing step was important as it enabled us to identify drivers with a significant impact on the target variable, and to drop those without any material effect (Punam et al., 2018; Cheriyan et al., 2018). Eventually a total of 11 variables (including three polynomial features) were found to have a significant impact on enterprise sales performance:

- Prod, Pemp, Temp, Temp2, LogTA, CoAge and Unemp at 1% significance level;
- CoAge2 and DigMkt at 5% significance level; and
- EntAge and EntAge2 at 10% significance level.

These identified variables were then used in building the machine learning models for OLS, LASSO, and ANNs, which were then evaluated to establish which one has the superior sales predictive accuracy.

#### ***Data partitioning, sales performance modelling, evaluation***

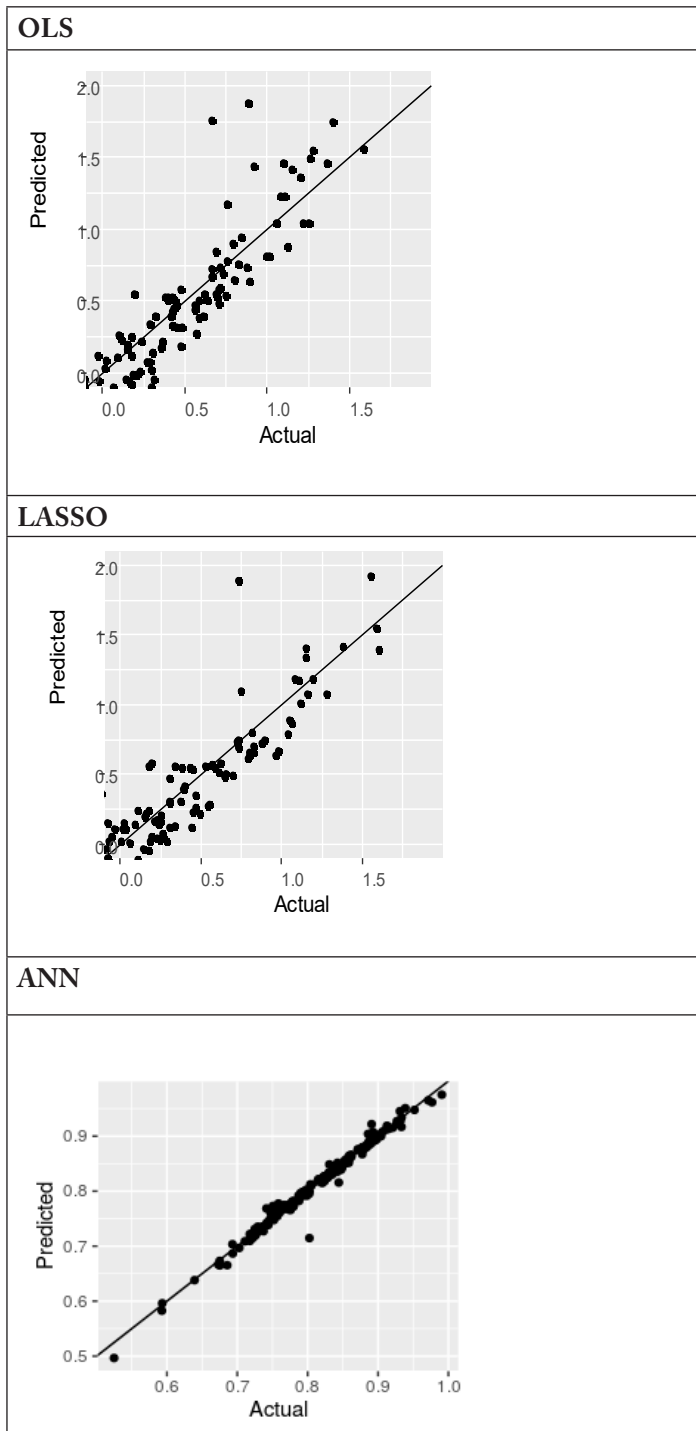
The next step was the dataset partitioning, which, as discussed above, is one of the critical elements in machine learning. For this study, following a related study (Delen et al., 2013), a 70:30 split ratio was used to generate the training and test datasets. Using these two datasets, three sales performance predictive models—one each for OLS, LASSO, and ANNs—were built and evaluated.

#### **4. Findings and analysis**

Figure 1 provides graphical representations of the predictive performance of each of the three algorithms—OLS, LASSO, and ANNs—on the test dataset. (The OLS algorithm was fit using the `plm` function in R. The LASSO algorithm was fit using the `glmnet` function in R, and 10-fold cross validation was performed to identify the optimal tuning parameter  $\lambda$ . The `neuralnet` function in R was used to fit the ANN algorithm, and the model with 2 neurons provided the best output and thus was used for further computations on the test dataset.)

The comparison shows that the OLS and LASSO algorithms' predictive performances are highly similar, and neither fits the data nearly as well as the ANN algorithm, which performs extremely well. Thus, the visualisations indicate that the ANN algorithm provides a more accurate sales predictions than do the other two algorithms.

Figure 1: Graphical representations of predictive performance



We also more formally evaluated each technique's predictive performance using five established model evaluation metrics: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute scaled error (MASE), and median absolute error (MDAE) (Casella et al., 2017; Hyndman & Koehler, 2006; Kolkman & Van Witteloostuijn, 2019; Muthukrishnan & Rohini, 2016; Punam et al., 2018; Tsoumakas, 2019). For each of the assessment metrics, the lower the value the better the algorithm's performance in predicting SMMEs' sales. The formal evaluation of the predictive models is presented in Table 2.

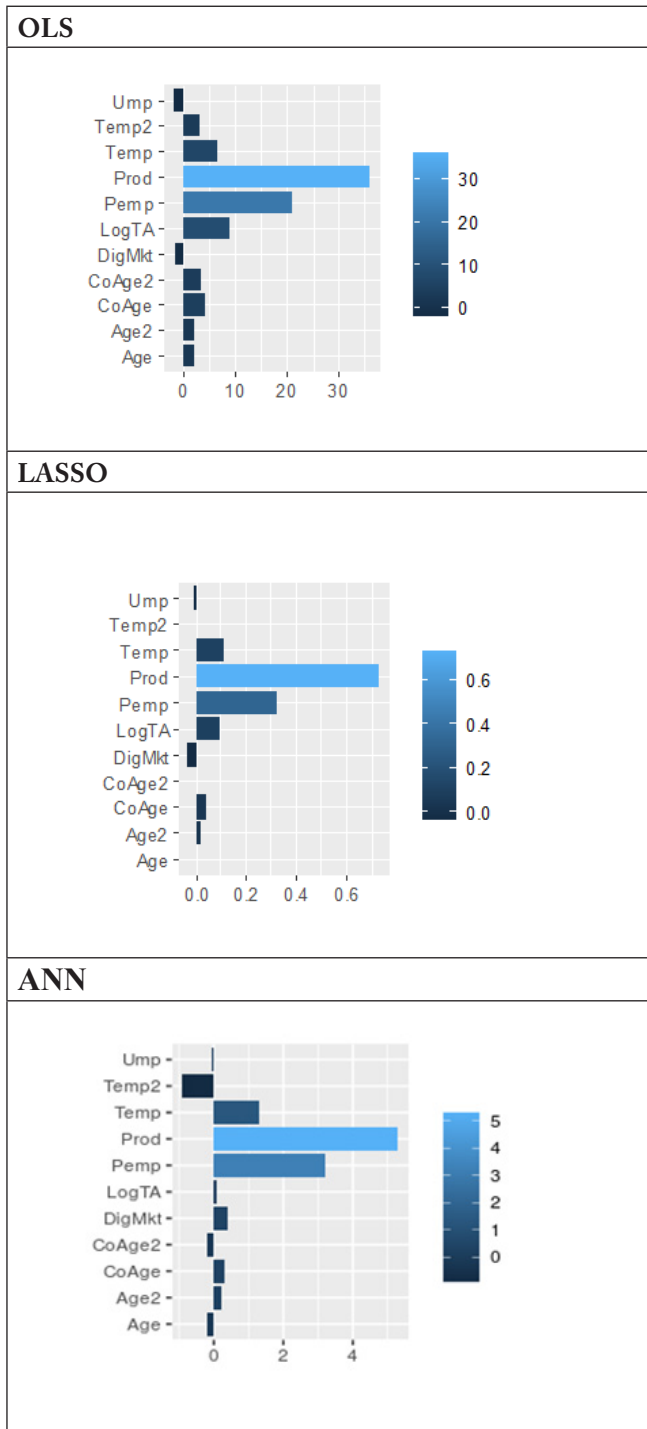
**Table 2: Formal evaluation of predictive performance**

	MSE	RMSE	MAE	MASE	MDAE
<b>OLS</b>	1.085967	1.042097	0.4517873	0.6902265	0.2088684
<b>LASSO</b>	0.08130493	0.2851402	0.1868477	0.2065746	0.1474003
<b>ANN</b>	0.000259	0.01611	0.0120	0.1389323	0.0072478

The assessment, as per Table 2, above shows that the ANN algorithm clearly outperforms other machine learning algorithms, as shown by the very low MSE, RMSE, MAE, MASE, and MADE values compared to the other two algorithms. The worst performing, as was expected, was OLS, with LASSO compared to the former showing some improvement on all the assessment metrics. Based on the assessment above and graphical analysis as per Figure 2, the ANN algorithm was thus selected as the best performing machine learning algorithm for sales predictive modelling.

Further to the evaluation of the predictive models, we also computed variable importance for each algorithm, as per Figure 3. From the graphical presentations it was clear across the three models that, generally, productivity and permanent workers are the two most important variables that positively influence SMMEs' sales performance. However, in respect of those variables which negatively influence sales performance, the results were mixed.

Figure 3: Variable importance





The ANN technique identified the excessive utilisation of temporary workers (Temp2) as negatively impacting on sales, while OLS indicated the opposite and LASSO showed no impact. Another feature which generated conflicting importance ratings was digital marketing, with OLS and LASSO highlighting it as having a negative impact on sales performance, while the ANN algorithm indicated that it had a positive effect.

The mixed findings show the importance of selecting the proper technique based on an objective criterion such as predictive accuracy. In this case, SMME owners would benefit most from exploiting ANN algorithms.

## 5. Conclusion and recommendations

The assessment found that the ANN approach was far superior to the other two machine learning approaches across all the assessment metrics, with the LASSO technique coming a distant second. The superior performance of the ANN algorithm, despite the inclusion of non-linear factors, shows the algorithm's versatility in identifying and incorporating functional relationships among variables in its predictive modelling process. This is in line with the assertion by Youn and Gu (2010) that, owing to their less restrictive assumptions when engaging with the dataset, ANNs tend to provide more accurate and reliable predictions than other algorithms.

More specifically, and in line with the other existing literature to date, this study provides stakeholders in the SMME sector with a basis for selecting ANNs to conduct sales predictive modelling and to inform strategic decision-making that can drive sustainable SMME growth. It is recommended that governments and other pertinent stakeholders develop and make available sales predictive applications powered by ANNs to manufacturing SMMEs in order to assist them in conducting predictions and developing data-driven plans. It is also recommended that future studies utilise larger datasets, covering periods longer than three years, to evaluate ANNs, and to compare ANNs' predictive performance with that of other complex techniques such as deep learning and support vector machines.

## References

- Adegbite, S., Ilori, M., Irefin, I. A., Aberejio, I., & Aderemi, H. O. S. (2007). Evaluation of the impact of entrepreneurial characteristics on the performance of small-scale manufacturing industries in Nigeria. *Journal of Asia Entrepreneurship and Sustainability*, 3(1), 1–21.
- Ahangar, R. G., Yahyazadehfar, M., & Pournaghshband, H. (2010). The comparison of methods artificial neural network with linear regression using specific variables for prediction stock price in Tehran Stock Exchange. *International Journal of Computer Science and Information Security (IJCSIS)*, 7(2), 38–46.
- Al-Ani, M. K. (2013). Effects of assets structure on the financial performance: Evidence from sultanate of Oman. In *11th EBES Conference proceedings*, 12–14 September, Ekaterinburg, Russia.

- Amran, N. A. (2011). The effect of owner's gender and age to firm performance: A review on Malaysian public listed family businesses. *Journal of Global Business and Economics*, 2(1), 104–116.
- Aziz, S., & Dowling, M. (2019). Machine learning and AI for risk management. In T. Lynn, J. Mooney, P. Rosati, & M. Cummins (Eds.), *Disrupting finance* (pp. 33–50). Palgrave Pivot. [https://doi.org/10.1007/978-3-030-02330-0\\_3](https://doi.org/10.1007/978-3-030-02330-0_3)
- Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481–485. <https://doi.org/10.1257/aer.p20151021>
- Bardasi, E., Sabarwal, S., & Terrell, K. (2011). How do female entrepreneurs perform? Evidence from three developing regions. *Small Business Economics*, 37(4), 417–471. <https://doi.org/10.1007/s11187-011-9374-z>
- Bauer, M. (2020). Machine learning framework for small and medium-sized enterprises. *SSRN 3532389*. <https://doi.org/10.2139/ssrn.3532389>
- Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, 53(2), 1051–1074. <https://doi.org/10.1007/s11135-018-0802-x>
- Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1), 133–153. <https://doi.org/10.1017/psrm.2014.7>
- Bellone, F., Musso, P., Nesta, L., & Quere, M. (2008). Market selection along the firm life cycle. *Industrial and Corporate Change*, 17(4), 753–777. <https://doi.org/10.1093/icc/dtn025>
- Bigsten, A., & Gebreyesus, M. (2007). The small, the young, and the productive: Determinants of manufacturing firm growth in Ethiopia. *Economic Development and Cultural Change*, 55(4), 813–840. <https://doi.org/10.1086/516767>
- Bureau for Economic Research. (n.d.). ABSA purchasing managers' index. <https://www.ber.ac.za/BER%20Documents/ABSA-PMI/?doctypeid=1066>
- Buyinza, F. (2011). *Performance and survival of Ugandan manufacturing firms in the context of the East African Community*. <https://ideas.repec.org/p/ags/eprcrs/150477.html>
- Camilleri, M. A. (2018). The SMEs' technology acceptance of digital media for stakeholder engagement. *Journal of Small Business and Enterprise Development*, 26(4), 504–521. <https://doi.org/10.1108/JSBED-02-2018-0042>
- Casella, G., Fienberg, S., & Olkin, I. (Eds.). (2017). *An introduction to statistical learning with applications in R*. Springer Texts in Statistics.
- Castelli, M., Dobрева, M., Henriques, R., & Vanneschi, L. (2020). Predicting days on market to optimize real estate sales strategy. *Complexity*, 2020. <https://doi.org/10.1155/2020/4603190>
- Chadwick, C., & Flinchbaugh, C. (2016). The effects of part-time workers on establishment financial performance. *Journal of Management*, 42(6), 1635–1662. <https://doi.org/10.1177/0149206313511116>

- Chen, J., De Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., ... Hoek, G. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment International*, 130. <https://doi.org/10.1016/j.envint.2019.104934>
- Cheriyian, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018). Intelligent sales prediction using machine learning techniques. In *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, Southend, UK, 16–17 August. <https://doi.org/10.1109/iCCECOME.2018.8659115>
- Clinebell, S. K., & Clinebell, J. M. (2007). Differences between part-time and full-time employees in the financial services industry. *Journal of Leadership & Organizational Studies*, 14(2), 157–167. <https://doi.org/10.1177/1071791907308053>
- Coad, A., Holm, J. R., Krafft, J., & Quatraro, F. (2018). Firm age and performance. *Journal of Evolutionary Economics*, 28(1), 1–11. <https://doi.org/10.1007/s00191-017-0532-6>
- Crane-Droesch, A. (2017). Semiparametric panel data models using neural networks. <https://arxiv.org/abs/1702.06512>
- Croda, R. M. C., Romero, D. E. G., & Morales, S.-O. C. (2019). Sales prediction through neural networks for a small dataset. *IJIMAI*, 5(4), 35–41. <https://doi.org/10.9781/ijimai.2018.04.003>
- Curran-Everett, D. (2018). Explorations in statistics: The log transformation. *Advances in Physiology Education*, 42(2), 343–347. <https://doi.org/10.1152/advan.00018.2018>
- Das, B., Nair, B., Reddy, V. K., & Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology*, 62(10), 1809–1822. <https://doi.org/10.1007/s00484-018-1583-6>
- De Kok, J., Ichou, A., & Verheul, I. (2010). New firm performance: Does the age of founders affect employment creation. *Zoetermeer: EIM Research Reports*, 12, 42–63.
- Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, 40(10), 3970–3983. <https://doi.org/10.1016/j.eswa.2013.01.012>
- Dod, H. S., & Sharma, R. (2010). Competing with business analytics: Research in progress. In D. N. Hart, & S. D. Gregor (Eds.), *Information systems foundations: Theory building in information systems* (pp. 239–249). ANU E Press.
- Droomer, M., & Bekker, J. (2020). Using machine learning to predict the next purchase date for an individual retail customer. *South African Journal of Industrial Engineering*, 31(3), 69–82. <https://doi.org/10.7166/31-3-2419>
- Egbunike, C. F., & Okerekeoti, C. U. (2018). Macroeconomic factors, firm characteristics and financial performance. *Asian Journal of Accounting Research*, 3(2), 142–168. <https://doi.org/10.1108/AJAR-09-2018-0029>
- Enkono, F. S., & Suresh, N. (2020). Application of machine learning classification to detect fraudulent e-wallet deposit notification SMSes. *The African Journal of Information and Communication*, 25, 1–12. <https://doi.org/10.23962/10539/29195>

- Essel, B. K. C., Adams, F., & Amankwah, K. (2019). Effect of entrepreneur, firm, and institutional characteristics on small-scale firm performance in Ghana. *Journal of Global Entrepreneurship Research*, 9(1), 55–75. <https://doi.org/10.1186/s40497-019-0178-y>
- Esteve-Pérez, S., & Mañez-Castillejo, J. A. (2006). The resource-based theory of the firm and firm survival. *Small Business Economics*, 30(3), 231–249. <https://doi.org/10.1007/s11187-006-9011-4>
- Farahani, D. S., Momeni, M., & Amiri, N. S. (2016). Car sales forecasting using artificial neural networks and analytical hierarchy process. In *Fifth International Conference on Data Analytics*, 9–13 October, Venice.
- Gepp, A., & Kumar, K. (2012). Business failure prediction using statistical techniques: A review. In K. Kumar, & A. Chaturvedi (Eds.), *Some recent developments in statistical theory and applications* (pp. 1–25). Brown Walker Press.
- Gholizadeh, P., Esmaceli, B., & Memarian, B. (2018). Evaluating the performance of machine learning algorithms on construction accidents: An application of ROC curves. In *Construction Research Congress 2018*, New Orleans. <https://doi.org/10.1061/9780784481288.002>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>
- Gupta, P. D., Guha, S., & Krishnaswami, S. S. (2013). Firm growth and its determinants. *Journal of Innovation and Entrepreneurship*, 2(1), 1–14. <https://doi.org/10.1186/2192-5372-2-15>
- Haataja, T. (2016). Sales forecasting in small and medium-sized enterprises. Master's thesis, Helsinki Metropolia University of Applied Sciences. <https://www.theseus.fi/handle/10024/106191>
- Halicioglu, F., & Yolac, S. (2015). Testing the impact of unemployment on self-employment: Evidence from OECD countries. *Procedia – Social and Behavioral Sciences*, 195, 10–17. <https://doi.org/10.1016/j.sbspro.2015.06.161>
- Harris, E. S. (1991). Tracking the economy with the purchasing managers' index. Federal Reserve Bank of New York, *Quarterly Review*, 16(3).
- Huggins, R., Prokop, D., & Thompson, P. (2017). Entrepreneurship and the determinants of firm survival within regions: Human capital, growth motivation and locational conditions. *Entrepreneurship & Regional Development*, 29(3–4), 357–389. <https://doi.org/10.1080/08985626.2016.1271830>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Jobs, C. G., & Gilfoil, D. M. (2014). A social media advertising adoption model for reallocation of traditional advertising budgets. *Academy of Marketing Studies Journal*, 18(1), 235–248.
- Kaunda, C. M. (2013). Entrepreneurial orientation, age of owner and small business performance in Johannesburg. Master's dissertation, University of the Witwatersrand, Johannesburg.

- Klapper, L., & Richmond, C. (2011). *Patterns of business creation, survival and growth: Evidence from Africa*. World Bank. <https://doi.org/10.1596/1813-9450-5828>
- Koenig, E. F. (2002). Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy. Federal Reserve Bank of Dallas, *Economic & Financial Policy Review*, 1(6), 1–14. <https://core.ac.uk/download/pdf/6971097.pdf>
- Kolkman, D., & Van Witteloostuijn, A. (2019). Data science in strategy: Machine learning and text analysis in the study of firm growth. *SSRN*. <https://doi.org/10.2139/ssrn.3457271>
- Krishna, D., Albinson, N., Chu, Y., & Burdis, J. (2017). *Managing algorithmic risks: Safeguarding the use of complex algorithms and machine learning*. Deloitte.
- Lantz, B. (2019). *Machine learning with R: Expert techniques for predictive modeling*. Packt Publishing.
- Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29–51. <https://doi.org/10.3390/risks7010029>
- Loderer, C. F., & Waelchli, U. (2010). Firm age and performance. *SSRN 1342248*. <https://doi.org/10.2139/ssrn.1342248>
- Maggina, A., & Tsaklanganos, A. (2012). Asset growth and firm performance evidence from Greece. *The International Journal of Business and Finance Research*, 6(2), 113–124.
- Melkumova, L., & Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering*, 201, 746–755. <https://doi.org/10.1016/j.proeng.2017.09.615>
- Merkel, G. D., Povinelli, R. J., & Brown, R. H. (2018). Short-term load forecasting of natural gas with deep neural network regression. *Energies*, 11(8), 2008. <https://www.mdpi.com/1996-1073/11/8/2008>
- Meroño-Cerdan, A. L., & Soto-Acosta, P. (2005). Examining e-business impact on firm performance through website analysis. *International Journal of Electronic Business*, 3(6), 583–598. <https://doi.org/10.1504/IJEB.2005.008537>
- Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: Algorithms and applications*. CRC Press. <https://doi.org/10.1201/9781315371658>
- Motoki, F. Y. S., & Gutierrez, C. E. C. (2015). Firm performance and business cycles: Implications for managerial accountability. *Applied Finance and Accounting*, 1(1), 47–59. <https://doi.org/10.11114/afa.v1i1.647>
- Muriithi, S. (2017). African small and medium enterprises (SMEs) contributions, challenges and solutions. *European Journal of Research and Reflection in Management Sciences*, 5(1), 36–48.
- Muthukrishnan, R., & Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. In *IEEE International Conference on Advances in Computer Applications (ICACA)*, 24 October 2016, Coimbatore, India. <https://doi.org/10.1109/ICACA.2016.7887916>
- Ndikum, P. (2020). Machine learning algorithms for financial asset price forecasting. *arXiv:2004.01504*. <https://arxiv.org/abs/2004.01504>
- Nghiep, N., & Al, C. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 313–336. <https://doi.org/10.1080/10835547.2001.12091068>



- Obaid, O. I., Mohammed, M. A., Ghani, M., Mostafa, A., & Taha, F. (2018). Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer. *International Journal of Engineering & Technology*, 7(4.36), 160–166.
- Panda, D. (2015). Growth determinants in small firms: Drawing evidence from the Indian agro-industry. *International Journal of Commerce Management*, 25(1), 52–66. <https://doi.org/10.1108/IJCoMA-12-2012-0080>
- Parsons, A. (2013). Using social media to reach consumers: A content analysis of official Facebook pages. *Academy of Marketing Studies Journal*, 17(2), 27.
- Pauka, K. (2015). *How does part-time work affect firm performance and innovation activity?* WWZ Working Paper No. 2015/05.
- Penpece, D., & Elma, O. E. (2014). Predicting sales revenue by using artificial neural network in grocery retailing industry: A case study in Turkey. *International Journal of Trade, Economics and Finance*, 5(5), 435–440. <https://doi.org/10.7763/IJTEF.2014.V5.411>
- Phillipson, J., Tiwasing, P., Gorton, M., Maioli, S., Newbery, R., & Turner, R. (2019). Shining a spotlight on small rural businesses: How does their performance compare with urban? *Journal of Rural Studies*, 68, 230–239. <https://doi.org/10.1016/j.jrurstud.2018.09.017>
- Punam, K., Pamula, R., & Jain, P. K. (2018). A two-level statistical model for big mart sales prediction. In *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*. <https://doi.org/10.1109/GUCON.2018.8675060>
- R Development Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ratnasena, N. H., Rich, D. C., Abraham, A. M., Cunha, L. L., & Morgan, S. L. (2021). Detection of magnetic audio tape degradation with neural networks and Lasso. *Journal of Chemometrics*, 35(1), e3194. <https://doi.org/10.1002/cem.3194>
- Rijkers, B., Söderbom, M., & Loening, J. L. (2010). A rural–urban comparison of manufacturing enterprise performance in Ethiopia. *World Development*, 38(9), 1278–1296. <https://doi.org/10.1016/j.worlddev.2010.02.010>
- Roca-Puig, V., Beltrán-Martín, I., & Cipres, M. S. (2012). Combined effect of human capital, temporary employment and organizational size on firm performance. *Personnel Review*, 41(1), 4–22. <https://doi.org/10.1108/00483481211189910>
- Ryll, L., & Seidens, S. (2019). Evaluating the performance of machine learning algorithms in financial market forecasting: A comprehensive survey. <https://arxiv.org/abs/1906.07786>
- Sekban, J. (2019). *Applying machine learning algorithms in sales prediction*. Master's thesis, Kadırhas University, Istanbul. <https://academicrepository.khas.edu.tr/handle/20.500.12469/2782>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107298019>
- Small Business Project. (2014). *Examining the challenges facing small businesses in South Africa*. Statistics South Africa. (2018). *Quarterly labour force survey: Quarter 2*. <http://www.statssa.gov.za/publications/P0211/P02112ndQuarter2018.pdf>



- Strandberg, R., & Låås, J. (2019). *A comparison between neural networks, Lasso regularized logistic regression, and gradient boosted trees in modeling binary sales*. Master's project, KTH Royal Institute of Technology, Stockholm.
- Te, Y.-F. (2018). Predicting the financial growth of small and medium-sized enterprises using web mining. Doctoral thesis, ETH Zurich. <https://www.research-collection.ethz.ch/handle/20.500.11850/309271>
- Thorsteinson, T. J. (2003). Job attitudes of part-time vs. full-time workers: A meta-analytic review. *Journal of Occupational and Organizational Psychology*, 76(2), 151–177. <https://doi.org/10.1348/096317903765913687>
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Statistical Methodology*, 73(3), 273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, 52(1), 441–447. <https://link.springer.com/article/10.1007/s10462-018-9637-z>
- Van Liebergen, B. (2017). Machine learning: A revolution in risk management and compliance? *Journal of Financial Transformation*, 45, 60–67. <https://ideas.repec.org/a/ris/jofitr/1592.html>
- Venishetty, S. V. (2019). *Machine learning approach for forecasting the sales of truck components*. Master's thesis, Blekinge Institute of Technology.
- Wang, P.-H., Lin, G.-H., & Wang, Y.-C. (2019). Application of neural networks to explore manufacturing sales prediction. *Applied Sciences*, 9(23), 5107. <https://doi.org/10.3390/app9235107>
- Youn, H., & Gu, Z. (2010). Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model. *International Journal of Hospitality Management*, 29(1), 120–127. <https://doi.org/10.1016/j.ijhm.2009.06.007>