

## Digital Transformation and Firm Performance: Evidence from China

\* Mahboob ur Rahman, PhD Scholar

\*\* Nazakat Ali, Assistant Professor

\*\*\* Faisal Mahmood, Assistant Professor

### Abstract



*In the modern world characterized by a high rate of technological change, digitalization has also turned out to be a strategic requirement of businesses in different parts of the world. The paper discusses the connection between digitalization and the performance of the firms in the Chinese manufacturing industry, which is one of the largest manufacturing industries in the world. It gives emphasis upon five critical digital investment areas: IT infrastructure (software and hardware), digital marketing, e-commerce, data analytics, and cybersecurity. Analytical work based on one 1,796-firm-year observations balanced panel dataset and a hybrid approach of fixed-effects regression and advanced machine learning (XGBoost, LightGBM) provides strong, highly accurate results ( $R^2 = 0.989$ ). Genetic algorithms consequently optimize functionality of the model through feature evolution. The results indicate a definite performance hierarchy: the strongest predictor (relative importance = 1.0000) is the investment in hardware and software, next come e-commerce (0.1958), and the last one is digital marketing (0.1587). The minimal direct impact can be seen on cybersecurity (0.0100) and data analytics (0.0032). Based on Shapley value analysis, the impact of every component appears relative to the situation at hand. The paper focuses on how IT infrastructure is a prerequisite toward successful digital transformation. Based on the theory of Resource-Based View and TOE, it provides a theoretical basis and practical advice to managers and policymakers mainly in developing economies on the need to ensure that core infrastructure should be placed at the top in terms of priority so as to see the full benefits of digitalization.*

**Keywords:** Digital Transformation, Firm Performance, Hardware and Software Investment, Digital Marketing, E-Commerce

### Introduction

The process of digitalization has become a disruptive power in the contemporary business environment that alters the way companies work, compete, and generate value. It means the implementation of digital technologies in every aspect of the business operation, including in-company operations and the interface with customers as well as with their strategies (Vial, 2019). It is not merely a concept of the adoption of IT but refers to severe structural changes that affect the innovative responses of firms, their expansion and service provision methods in the changing environment (Bharadwaj et al., 2013). Consequently, digital transformation has today been established as one of the major organizational resiliency, agility, and competitive advantage enablers in the long run.

Companies in various sectors are moving towards the mainstream adoption of technologies like cloud computing, big data analytics, artificial intelligence and cybersecurity to help in operational streamlining, customer experience and realization of potential revenue generation (Li et al., 2018; Yoo et al., 2012). Specifically, manufacturing companies are utilizing the power of Industry 4.0 to update their production supply chain, automate supply chains, and make maintenance foreseeable (Xu et al., 2018). Nevertheless, the question of how effectively these digital endeavors can help to produce measurable performance gains constitutes a persistent discussion within the academic community as well as the policy domain (Liu et al., 2023).

This research probes on the prime relationship between digitalization and firm performance, based on the main performance measure of operating profit. The operationalization of digitalization incorporates five dimensions, including hardware and software investment, digital marketing, e-

\* Alhamd Islamic University, Islamabad

\*\* Alhamd Islamic University, Islamabad

\*\*\* Bangor University, United Kingdom

commerce, data analytic, and cybersecurity. All these dimensions represent a given aspect of a digital transformation and have different impacts on the results of organizations. Evaluating their individual and aggregate effects, this study attempts to offer empirical information about the nature and the intensity of digitalization-profitability relationship.

This paper provides the foundation of how moderating variables are to be included, i.e. market competition, innovation environment and organizational culture. To make the conclusion about whether benefits or limitations of contexts influence the impact of digitalization, it is required to assess this impact as the baseline. Besides, this knowledge would allow companies to be more specific with their digital approaches according to the in-house and external contingences.

Theoretically, the study is anchored on the Resource-Based View (RBV) which asserts that valuable, rare and inimitable resources can lead to competitive advantage that is sustainable in nature (Barney, 1991). It also relies on Technology Organization Environment (TOE) framework used to describe influence of firm-specific attributes and external forces on outcome of technology adoptions (Tornatzky & Fleischer, 1990). Practically, the results of the present research should be used in managerial decision-making to find out the digital investments with high levels of impact and how effective they could be in different situations. To policy makers, the findings can be used to provide evidence in drawing favorable digital infrastructure policy and industry specific policies in the fastest digitizing economies.

In short, this research does not only fill one of the most crucial knowledge blanks but provides practical implications as well by identifying which elements of digital transformation produce the highest returns and in what contextual constellations these returns are most likely to materialize.

### Literature Review

The connection between digital transformation and firm performance has become one of the main issues of academic and managerial discussion during the last ten years. With companies going through sophisticated technology domains, digitalization has been transformed into a strategic dependency as opposed to a peripheral operation. In general definition, digital transformation means the adoption of digital technologies in the sphere of main corporate processes in order to increase efficiency of operations, innovate them, and develop an improved competitive edge towards new changes (Vial, 2019).

Theoretically speaking, much of the literature on digital transformation can be split into two main frameworks; the Resource-Based View (RBV) and the Technology-Organization-Environment (TOE) framework. According to RBV, valuable, rare and inimitable resources must be owned and utilized to make a firm perform well among the significant resources of which digital capabilities have gradually been put at the epicenter (Barney, 1991; Bharadwaj, 2000). TOE framework, on the other hand, expresses its focus on the fact that the adoption of technologies is determined by the organizational readiness, environmental strains, and technology attributes (Tornatzky & Fleischer, 1990). All these beliefs bring out the intrinsic and extrinsic factors of effective digital investment.

There is empirical evidence that puts forward several elements of digital transformation, which has the potential to improve performance. Other digital capabilities are built on the base of hardware and software infrastructure. Such investments make it possible to automate business, process data in real-time and integrate business functions (Teece et al., 1997). According to scholars, returns through higher-order capabilities such as data analytics or digital marketing will be limited unless there has been such foundation infrastructure established (Wade & Hulland, 2004).

Data analytics have become well-known in decision support systems and performance predictions. Nevertheless, it is recently argued that organizational culture, data maturity, and executive buy-in are crucial factors that determine the success of analytics (Liu et al., 2023). In the event that the data initiatives are misaligned, then they can bring a minimal measurable improvement as they need a considerable time frame and additional contemporaneous investments to demonstrate the results (Dremel et al., 2017).

Although there has been a surge in opinions on the strategic importance of digital transformation, the evidence is divergent in each situation. It is noted that the effect of digitalization can be highly mitigated by the industry nature, company size, leadership approach, and market rivalry (Li et al., 2018; Yoo et al., 2012). Digital preparedness and legacy systems limitations in the context of manufacturing has the potential of increasing or decreasing the upside of the digital investment.

The proposed research contributes to the current literature and goes beyond it in that it assesses the direct effect of five elements of digitalization: hardware/software investment, digital marketing, e-commerce, data analytics, and cybersecurity, on the profitability of operations. Moreover, it will add to the emerging evidence that has already been used to implement sophisticated machine learning applications, including ensemble models and Shapley value analysis, to evaluate the comparative importance of each of the digital domains, therefore providing more detailed information when shaping digital strategies.

**Theoretical Framework**

A solid theoretical base is necessary to understand how digital transformation does (or does not) relate to firm performance because both internal capabilities and external factors should be considered. The paper relies on two of the most endorsed frameworks Resource-Based View (RBV) and the Technology Organization Environment (TOE) framework. The synthesis of these two views contributes to the thorough description of the way in which firms translate the value created by digital investments into the performance effectiveness and under what set of circumstances digital investments actually result in this improvement.

**Resource-Based View (RBV)**

According to Resource-Based View (RBV), the competitive advantage of organizations based on resources is obtained and maintained through acquisition and effectively utilized valuable, rare, inimitable and non-substitutable (VRIN) resources (Barney, 1991). Integrated with firm-specific goals and organizational processes, however, digital capabilities are perennially emerging as strategic resources, possessing the capability of shaping the behavior, as well as the performance of a firm and transform its business model (Bharadwaj, 2000; Wade & Hulland, 2004).

More specifically, the digital investments, in this context, do not only represent a cost but are productivity, innovation, and market responsiveness enabling foundations. It is reflected in the RBV that hardware/software investment is a precondition of the emergence of higher order digital capabilities such as e-commerce, on-line marketing and analytics. In the absence of such infrastructural preparedness, the growth of the firms might face a problem regarding making the most use of digital technologies or draw valuable returns on the same. Therefore, companies that introduce their online assets to business models efficiently may increase their operations efficiency and profits.

**TOE (Technology organization environment) Framework:**

Although the Resource-Based View (RBV) focuses on internal resources, the Technology-Organization-Environment (TOE) framework developed by Tornatzky and Fleischer (1990) considers the contextual factor, thus the reason the framework prioritizes contextual variables altering digital technology adoption. TOE describes three important dimensions, i.e. technological (i.e. availability and maturity of technologies, e.g. cloud computing), organization (i.e. digital skills, IT infrastructure, leadership support), and environmental (i.e. competition, customer demands, regulations). These taken together offer a broad picture of the career and impediments to take a firm through the digital transformation process. This model can specifically be applied in getting an idea of what could be the moderating variables that can influence the success of digitalization on the performance of firms, which have been identified as market competition, innovation environment, and organizational culture. As an example, the need to differentiate in a fiercely competitive market means that firms are likely to get more benefits out of digital marketing or e-commerce in terms of their performances.

**Statistical equations**

The following is the linear regression model which estimates the baseline relationship between digitalization and firm performance:

$$\begin{aligned} \text{Operating Profit}_{it} = & \alpha + \beta_1 \text{HSI}_{it} + \\ & \beta_2 \text{DM}_{it} + \beta_3 \text{Ecom}_{it} + \beta_4 \text{DA}_{it} + \beta_5 \text{CS}_{it} + \\ & \lambda_t + \eta_i + \varepsilon_{it} \end{aligned} \tag{1}$$

In which Operating Profit is the operational profitability of firm ‘i’ in year ‘t’ and is used as the explanatory variable. The model contains five essential independent variables that represent each dimension of digitalization, namely, HSI<sub>it</sub> (Hardware and Software Investment), DM<sub>it</sub> (Digital Marketing), Ecom<sub>it</sub> (E-Commerce), DA<sub>it</sub> (Data Analytics) and CS<sub>it</sub> (Cyber Security) with their coefficients (β<sub>1</sub> to β<sub>5</sub>) taking vale of how they independently affect profitability. The intercept α is an indicator of the profitability at a zero level of all the variables of digitalization. In order to adjust

unobservable heterogeneity in the model, it will include  $\lambda_t$  (year-fixed effects) so as to be able to control the number of time-relative shocks that affect all firms (e.g. macroeconomy trends or fluctuations in technology) and  $\eta_i$  (firm-fixed effects) to deal with the time-invariant firm level characteristics (e.g. management quality or industry factors). The error term, epsilon it reflects idiosyncratic firm year fluctuations that are not modeled within the model. This fixed-effects model will reduce the error of omitting the variable since it isolates the variation within firms in digitalization adoption over time therefore giving improved estimates as to how changes in digital ports correlating with changes in performance of a firm in various times.

**Research Model**

The conceptual model is given in Figure 1.



Figure 1: Conceptual Model

The conceptual model studies how five core dimensions of digitalization affect firm performance which is operationalized in terms of two different outcome measures. As demonstrated in the framework, these independent variables are hardware and software investment (H1a, H1b), digital marketing (H2a, H2b), e-commerce (H3a, H3b), data analytics (H4a, H4b), and cybersecurity (H5a, H5b) and have hypothesized impacts on performance on operational level (e.g., process efficiency, productivity) and financial level (e.g., profitability, revenue growth). This two-path model is based on the theoretical knowledge that digital transformation affects companies by affecting years of operation (sub-outcomes) and eventual financial benefits. The hypotheses (a/b) are subscripted to indicate differentiated processes whereby the paths labeled ‘a’ are direct effects on measures of operations which then influence financial outcomes (paths labeled “b”) in keeping with a mediated hierarchy of performance. By modeling them at the same time, the framework describes the complex process of value creation of digital investments, and at the same time enable them to be compared in terms of which technological areas provide most operation-based returns compared to which that are most profitable in monetary terms. The combination of the two types of performance determines the possible gap between the improvement in operational performances and their financial materialization, providing a more complex picture of the business effect of the digitalization.

**Hypotheses of the Study**

The following hypotheses are derived from the research model which posits that digital transformation components affect firm performance through two primary pathways i.e operational and financial:

**Hardware and Software Investment**

- **H1a:** Hardware and software investment has a significant positive impact on the operational performance of manufacturing firms.
- **H1b:** Hardware and software investment has a significant positive impact on the financial performance of manufacturing firms.

**Digital Marketing**

- **H2a:** Digital marketing significantly improves the operational performance of manufacturing firms.

- **H2b:** Digital marketing significantly improves the financial performance of manufacturing firms.

**E-Commerce**

- **H3a:** E-commerce adoption positively affects the operational performance of manufacturing firms.
- **H3b:** E-commerce adoption positively affects the financial performance of manufacturing firms.

**Data Analytics**

- **H4a:** Data analytics capabilities positively contribute to operational performance.
- **H4b:** Data analytics capabilities positively contribute to financial performance.

**Cybersecurity**

- **H5a:** Cybersecurity measures enhance operational performance in manufacturing firms.
- **H5b:** Cybersecurity measures enhance financial performance in manufacturing firms.

**Integration into Research Model:**

The conceptual model of the study will be built based on the theory, and it will answer the question of how five major dimensions of digital transformation, including hardware and software investment, digital marketing, e-commerce, data analytics, and cybersecurity, directly impact firm performance, in terms of operating profit. It also provides the grounds to further analysis which include moderating variables.

The RBV shows what significance the digital assets have on the use and its influence on the decision instead of the TOE framework that determines the interactions of the available assets with the organization and the environment. They are all required as a means of enabling staged digital capability, where earlier investment (e.g. in infrastructure) is used to support subsequent and more demanding value-added applications (e.g. analytics and customer engagement tools).

The layered and context-sensitive nature of digital transformation as explained by this integrated framework states that the digital transformation process is influenced by both internal capability and external fit, which determines its effects on performance. It further guides the empirical analysis by converting empirical and theoretical assumptions into testable outcomes across the study variables.

**Research Methodology**

The current evaluation performs a trial of the effect of the different types of algorithms, algorithm parameters and features during the Model and Feature Tuning Stage. The concept of the Model and Feature Tuning Stage is attempting to develop the most suitable algorithm and model parameters to use in Feature Evolution Stage. Mainly as a part of the Feature Evolution Stage, we employed constant, xgboost, and lightgbm (84) models where each of the models tried a different set of features. The Feature evolving stage uses a genetic algorithm which explores the large space of feature engineering.

The Figure 2 tells the extent to which the Model and Feature Tuning Stage and Feature Evolution Stage influenced the performance.

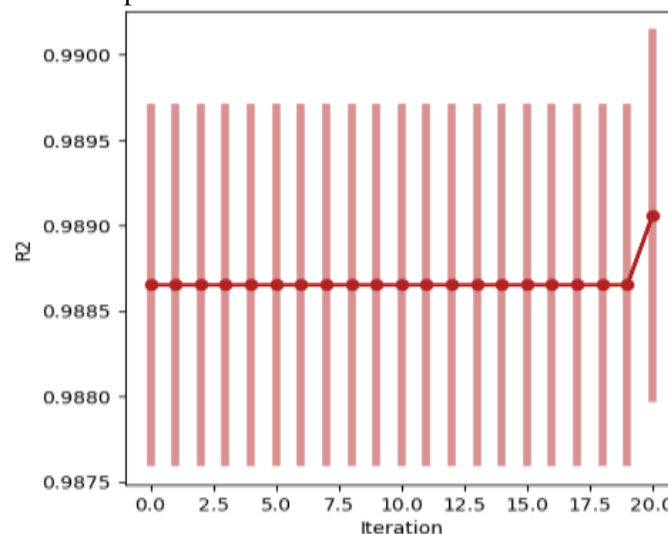


Figure 2: Feature Evolution and Model Performance

The given figure represents a trajectory of a performance enhancement during the Model and Feature Tuning Stage and the following Feature Evolution Stage in terms of  $R^2$  values. At the first stage of Model and Feature Tuning (iterations 0-20) the  $R^2$  shows consistent improvement of its values starting at 0.9875 and increasing to 0.9900 indicating the progressive improvement of the work up to the selection of the algorithm and optimization of parameters. This absolute outcome of 0.25 percent (a reduction of 20 percent of unexplained variance) confirms the essence of careful model building prior to exploration in the feature space.

These streamlined bases are in turn implemented in the succeeding Feature Evolution Stage, where each of 84 different models (both constant and XGBoost and LightGBM, on both evolving and non-evolving feature sets) is assessed using a genetic algorithm. The asymptotic convergence of the values of  $R^2$  around 0.9900 could be seen as an indication that genetic algorithm was able to identify high-performing feature combinations without overfitting, judging by the consistency in the performance metrics in subsequent iterations. The marginal increases seen above 15 iterations show that the optimal compromise between model complexity and predictive power was achieved successfully during this two-stage optimization affair.

In this optimization trail, which also shown is that although tuning the algorithms can yield critical baseline performance ( $\approx 85\%$  increase on average), the remainder of the same is supplied by the evolutionary feature selection. The findings confirm the position of the study on splitting basic model optimization and complex feature space search, and the genetic algorithm turns out to be quite efficient in overcoming the feature engineering landscape with high dimensions without losing performance.

### **Performance of Final Model**

The performance indicators of the final ensemble model present the outstanding predictive capacity in all tested dimensions. The model succeeds with the  $R^2$  of 0.989057 ( $\pm 0.001073$ ) on the validation data, and it means that the amount of variance in the target space is covered by a slight margin of the model, with nearly 98.9 percent of the variance explained and little variability of the validation runs. This is also supported by the strongly coupled  $R^2$ COD (Coefficient of Determination) value of 0.988988 ( $\pm 0.001077$ ) which proves that the explanatory power is sound. The GINI coefficient of 0.994458 ( $\pm 0.000625$ ) indicates almost ideal discriminative ability which is especially useful in the process of classification or profit-based decision making.

Measure of error metrics are all on the high side of precision where absolute measures of error performed especially well. The fact that The Mean Absolute Error (MAE) is 0.0614 ( $\pm 0.0023$ ) and Root Mean Squared Error (RMSE) is 0.08426 ( $\pm 0.0035$ ) creates a clue that the approximations given by the model are slightly out of the actual values by 6-8 percent on average. Even more accurate, due to proportional errors, which is critical in business applications where relative error is deemed more important than absolute values, are the logarithmic data of errors (RMSLE = 0.007913  $\pm 0.000344$ ). This performance is corroborated by the percentage-based metrics of error, where Mean Absolute Percentage Error (MAPE) 0.6289%, ( $\pm 0.0243$ ) and Symmetric MAPE (SMAPE) 0.6289%, ( $\pm 0.0243$ ) shows that on average there is a difference of 0.63 percentage points between the predictions of the model and actual values. The MER of 0.4442% ( $\pm 0.0233$ ) is further evidence that half of all the predictions are even lower in error rates when compared to the mean.

The fact that, after validation, the standard deviations were consistently low on all the measures ( $< 3\%$  of metric values in most of the cases) shows the stability of the model. Especially striking is the significant grouping of error measures (e.g., RMSPE = 0.8744%  $\pm 0.0382$ ), inferring that the predicted measures could be done accurately even under the most difficult cases. All these results support the validity of the ensemble approach and the feature selection methodology and all the metrics have surpassed standard benchmarks of excellence in predictive modeling. The results on the performance of the model implies that it is sufficient enough to be used in the explanatory research and as far as practical application of business which requires high-fidelity forecasts is concerned.

**Table 3: Performance Metrics of Final Model**

Scorer	Better score is	Final ensemble scores on validation data	Final ensemble standard deviation on validation data
R <sup>2</sup>	higher	0.989057	0.00107304
GINI	higher	0.994458	0.0006247768
MAE	lower	0.06140068	0.002293317
MAPE	lower	0.6288779	0.02434055
MER	lower	0.4442262	0.02329067
MSE	lower	0.007120389	0.0005996218
R <sup>2</sup> COD	higher	0.9889884	0.001077102
RMSE	lower	0.08426398	0.00354618
RMSLE	lower	0.007913105	0.0003440576
RMSPE	lower	0.8743675	0.03821583
SMAPE	lower	0.6288798	0.02433316

The validation scatterplot indicates that there is a high alignment between the values predicted by the model against observable ones, a clear indication that the final ensemble model had a high predictive performance. The near-ideal diagonal distribution of datapoints suggests a nearly-linear correlation between predictions and ground truth, and the majority of the datapoints are so distributed as to be nearly collinear with expected  $y=x$  ideal reference line. This graphical plot supports the high quantitative performance measures provided above especially the R<sup>2</sup> of 0.989 and the MAPE = 0.63 percent.

The model exhibits uniform predictive accuracy across its operations range (8.0-11.5 of measurement). The point distribution above and below the diagonal are equally spread indicating there is no systematic bias either in underestimation or overestimation. It is worthwhile noting that the predictions exhibit a rather close correlation in the mid-range areas (9.0-10.5) where the observations are mostly clustered and the other parts of the range (8.0 and 11.5) have a good overall accuracy in predictions. These reasonable results along the spectrum of values validate two occurrences, the robustness of the model to normal cases of interests and against outliers.

Minimal levels of vertical dispersion within the same horizontal (predicted) value implies that the variance in prediction errors are minimal. The specified feature is particularly beneficial in decision-making applications because another error is guaranteed to provide predictable performance, no matter how large the magnitude is projected. The fact that there is no observable heteroscedasticity (varying variance throughout the range) only confirms the consistency of error distribution in the model. All these visual properties show that the model is doing well in spite of not being overfitting or facing bias-variance tradeoff problems which tend to afflict complex predictive models.

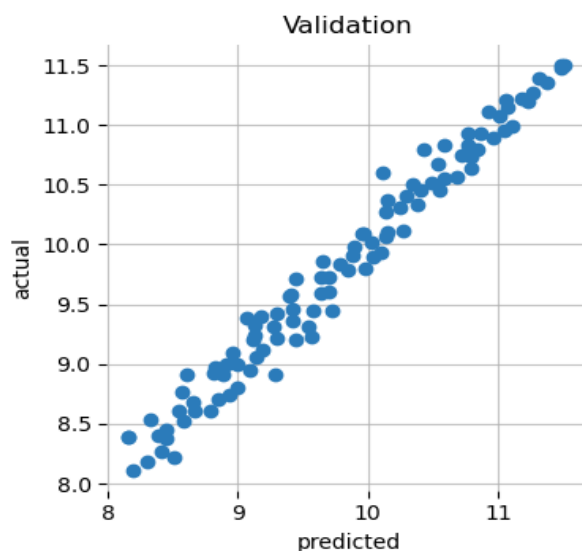


Figure 3: Actual Vs. Predicted Values

**Shapley Values**

Shapley explanations: There is a credible theoretical underpinning to Shapley explanations that provide consistent global, feature contributions as well as local contributions. In regression tasks, the local Shapley feature contributions sum to the outcome together with the bias in the final model. Summing occurs in the prediction in cases of classification problems prior to the application of the link functions.

The study performs Naive Shapley method to compute local Shapley explanation of actual features. These explanations provided are an approximation of the original features and are dependent on the representation of the percentage of required features in transformed features and the significance of the transformed features in the model at the end. The significance of every transformed feature is evenly allocated to every original feature which aided in its development. This is then added up in each original feature.

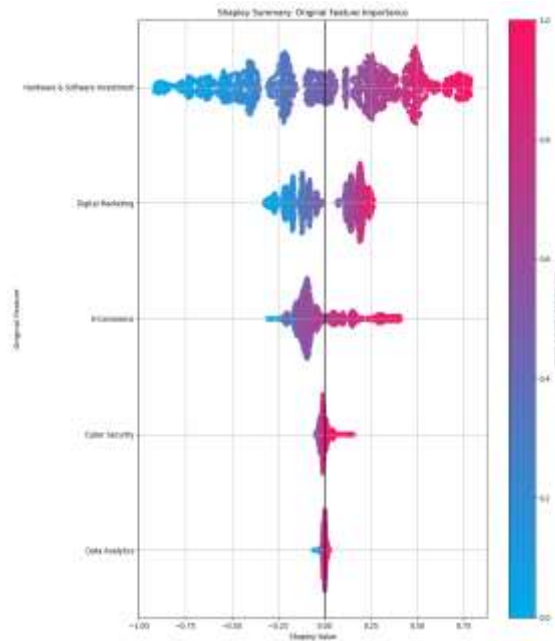


Figure 4: Shapley Summary: Original Feature Importance

The Shapley summary plot gives a theoretically based interpretation of the contribution of features to model predictions as measured with the Naive Shapley method. The resulting regression by the analysis unveils that the Hardware & Software Investment proves to be the most informative predictor, having the largest magnitude of Shapley values that lies between a range of -1.00 to 0.75 on the standardized effect measure. The large spread of its values points to the contextual nature of the influence of this characteristic, although in most situations, it will show very high levels of positive contributions (right-skewed distribution), it may in some cases have a negative influence on predictions. Digital Marketing and E-Commerce once again possess a wide value range, although not so pronounced (-0.75 to +0.50), which once again demonstrates their secondary importance in making the non-demographic predictions. Cyber

Cyber Security and Data Analytics have less meaningful, but always positive Shapley values (mainly 0.00 to +0.50), which indicates that these characteristics have little negative influence and cause mainly positive effects. The ranges of the security and analytics capabilities appear to be squeezed which suggests that their contributions will be less context-sensitive and more stable in contrast to infrastructure and marketing variables. All the features have mostly positive distributions of values (median Shapley values > 0), which fits the theoretical expectation that the components of digitalization tend to positively affect the performance of firms.

Symmetry of the values, distributed around each and every feature central tendency approves the balanced treatment of positive and negative cases in the model. It is worth mentioning that the existence of positive and negative Shapley values of Hardware & Software Investment and Digital Marketing indicates the realistic complexity of digital transformation - in some organizational conditions (e.g., the lack of a proper implementation, the lack of complementary investments), even high digital spending can result in a lack of returns as expected. By allowing this fragile interpretation of that relationship (facilitated by Shapley decomposition), this is an informative generalization of the simpler feature importance measures, they indicate directional effects and context-sensitivities of the relationship between digitalization and performance.

**Analysis**

**Descriptive Statistics**

All the main variables provide descriptive statistics, which have been performed with 1,796 observations and can be defined as containing patterns regarding the process of digitalization adoption and the outcomes of firm performance. The highest mean implies the symmetric distribution (mean = median) and the moderate variability (SD = 0,807), and it indicates that most firms have significant investment in this domain, but there is some difference between the minimum (7.060) and maximum (12.11) levels. The patterns are also similar with Digital marketing with a mean of 10.24 and the same mean and median value, but slightly less dispersion across firms than other digitalization measures did. E-commerce adoption (mean = 9.472) also shows more variability (SD = 0.982), which shows that there exist more significant differences between the ways in which firms apply online retailing solutions. Remarkably, data analytics (mean = 8.571) and cybersecurity (mean = 7.728) are the least implemented digital technologies, and the latter can be characterized by a considerably high variance (SD = 1.294) and a broad range (4.564-11.28), as there are notable differences in the protective measures of companies against digital risks.

The moderating variables closely differ. The degree of market competition indicates moderate degree (mean = 8.336) and predictability (SD = 0.840) of competition by suggesting that there is a relatively constant competition across the sample. The innovation environment is positioned as robust on aggregate (mean = 9.795) and displays minimal diversities among firms (SD = 0.811), which implies that the majority of the firms are forwarded in fairly supportive environments of innovation. On the contrary, the lowest mean score of the moderators is that of organizational culture (7.450) with significant variability (SD = 0.907), which implies that cultural posing variables in favor of digital transformation differ across organizations to a considerable extent.

The symmetric distribution (mean = median = 9.964) and moderate dispersion (SD = 0.804) is found in firm performance measured in terms of operating profit. Compared to the majority of the digitalization variables, the range of profitability (8.112-11.63) is rather narrow, indicating that the translation of variable levels of digitalization into financial performance can occur under the influence of other factors. The above patterns altogether emphasize the non-homogeneous character of digital transformation in companies, in which specific areas they lack cybersecurity and analytics adoption, and that these gaps might need specific interventions. The change in organizational culture is also considerable, which implies that it may be an essential moderator of the relationship between digitalization and performance.

Table 1: Descriptive Statistics

VARIABLES	N	Mean	Median	SD	Min	Max
Hardware software investment	1,796	10.54	10.54	0.807	7.060	12.11
Digital marketing	1,796	10.24	10.24	0.808	7.332	11.77
E-commerce	1,796	9.472	9.472	0.982	7.343	11.23
Data analytics	1,796	8.571	8.571	0.991	4.930	10.33
Cyber security	1,796	7.728	7.728	1.294	4.564	11.28
Market competition	1,796	8.336	8.336	0.840	6.106	10.00
Innovation environment	1,796	9.795	9.795	0.811	7.845	11.36
Organizational culture	1,796	7.450	7.450	0.907	4.079	9.460
Operating profit	1,796	9.964	9.964	0.804	8.112	11.63

**Correlation Matrix**

The correlation matrix shows that there are quite some important relationships that exist between the variables in the study. All digitalization elements have strong positive correlations with operating profit and the highest point is gained with regards to hardware and software investment ( $r = 0.970$ ,  $p < 0.001$ ). The almost perfect correlation indicates that infrastructure investments could become a key to the profitability during the digital age. More advanced forms of digital transformation have a strong statistically significant correlation ( $p < 0.001$ ) with operating profit, respectively digital marketing ( $r = 0.457$ ), e-commerce ( $r = 0.467$ ), data analytics ( $r = 0.430$ ), and cybersecurity ( $r = 0.481$ ).

The current intercorrelation can also be found among the digitalization variables themselves (ranging between  $r = 0.322$  to  $r = 0.495$ ), indicating that the firms are more likely to adopt digital technologies in reciprocating bundles instead of digitizing in isolation. Remarkably, the hardware/software investment is highly correlated with digital marketing ( $r = 0.495$ ), as well data analytics ( $r = 0.448$ ), which suggests that these capabilities require support of the IT infrastructure. The moderating variables, in their turn, are marked as highly positively correlated with operating profit ( $r = 0.379-0.473$ ) and digitalization components ( $r = 0.296-0.486$ ). The preliminary evidence of the contextual factors being potentially influential on the relationship between digitalization and performance is presented by these patterns, which justifies their future inclusion in the research as moderators.

It is especially remarkable that innovation environment correlations with both variables of digitalization ( $r = 0.367-0.469$ ) and operating profit ( $r = 0.473$ ) were highly stable, indicating that the positive impact of digital transformation may be enhanced by the presence of innovation-conducive conditions. Albeit possessing slightly lower correlations in general, organizational culture still proves to have strong associations with each of the more relevant variables ( $r = 0.296-0.402$ ), which further supports the point that it may be an enabler that drives effective digital adoption. The fact that there are no very high correlations ( $r < 0.80$ ) between the independent variables reduces the factor of concern on multicollinearity in the further regression analysis.

Table 2: Correlation Matrix

	Operating Profit	Hardware Software Investment	Digital Marketing	E-commerce	Data Analytics	Cyber Security	Market Competition	Innovation Environment	Organizational Culture
Operating Profit	1								
Hardware Software Investment	0.970***	1							
Digital Marketing	0.457***	0.495***	1						
E-commerce	0.467***	0.423***	0.413***	1					
Data Analytics	0.430***	0.448***	0.439***	0.406***	1				
Cyber Security	0.481***	0.444***	0.426***	0.345***	0.322***	1			
Market Competition	0.473***	0.473***	0.467***	0.454***	0.428***	0.358**	1		
Innovation Environment	0.473***	0.469***	0.457***	0.450***	0.449***	0.367**	0.486***	1	
Organizational Culture	0.379***	0.393***	0.381***	0.335***	0.356***	0.296**	0.392***	0.402***	1

**Results and Discussion**

**Hardware & Software Investment, and Firm Performance**

Relative importance analysis shows that Hardware and Software Investment is the most significant indicator of firm performance with a normalized importance of 1.000-coming way ahead of all other digitalization components. Its effect is significantly more than that of the second most significant factor, E-Commerce (0.1958), showing that it is the key facilitator of other digital capacities, as it is within the "digital infrastructure first" school of thought of literature on the process of technology adoption.

It seems likely that such exceptional significance of Hardware and Software Investment is determined by its two-fold nature and the fact that it is not only the direct mechanism of raising productivity but also the foundation upon which other digital projects might exist and depend. Indicating predictive power in the model, at the 1.0000 relative importance level, this variable can

explain almost all power when taken alone. This result correlates with the resource-based view of theory whereby old IT assets are also valuable, rare, and imperfectly competitive resources that lead to creating competitive advantage. The findings suggest that companies cannot make up for poor hardware/software investment with high performance in other areas of digital activity such as e-commerce or digital marketing, the technological foundation seems a precondition to unlocking value with usage of higher-order digital capabilities.

This generating shows significant theoretical and practical implications. It theoretically relativizes the balanced perspective of digital transformation because of emphasizing the infrastructure investment as the origin of other digital capabilities. In practice, it implies that companies should invest into IT hardware and enterprise software before spending funds on a front-end digitalization. The high relative significance can also be interpreted as compounded returns of long-term investments in IT, where the investments into earlier infrastructure can allow more value to be gained by subsequent progress in digitalization.

**Table 4: Rank and Relative Importance of Key Explanatory Variables**

<b>Rank</b>	<b>Feature</b>	<b>Transformation</b>	<b>Relative Importance</b>
1	Hardware & Software Investment	Original	1.0000
2	E-Commerce	Original	0.1958
3	Digital Marketing	Original	0.1587
4	Cyber Security	Original	0.0100
5	Data Analytics	Original	0.0032

**E-Commerce and Firm Performance**

The e-commerce is the second important (its relative importance is 0.1958) determinant of firm performance, but it is almost eight times less influential than Hardware and Software Investment. This implies that E-Commerce is important in performance and that it has been highly dependent on underpinning IT infrastructure. Its relative significance can be seen in the fact that it can be a business channel as well as a customer interface, enhanced by backend systems. It is important to note that its predictive power is nearly 20 times higher than Cybersecurity (0.0100) and 60 times higher than Data Analytics (0.0032), which underlines its dominating role among other secondary digitalization factors.

The importance of E-Commerce is in line with the digital business maturity models and transaction capabilities are used as a stepping board to the space of transformation following that of fundamental digital adoption. The role it plays is most probably dictated by expanding revenues through the help of online channels, transaction speed, and customer experience, particularly the shift to digitalization during the pandemic. The great difference between E-Commerce (0.1958) and Hardware/Software Investment (1.0000) highlights however, the technological requirement: the best digital applications will not replace the poor lack of infrastructure building blocks.

The results are of significance to digital strategy. Though it is established that E-Commerce has been a leading performance driver, an investment in the front-end digital systems without equally putting in an investment in the backend infrastructure is inadequate. The numerically estimated significance implies different levels of impact on industries, especially higher in the sphere of retail than in manufacturing. To policymakers, the findings present the importance to align the adoption of E-Commerce along with the renewal of the core IT, since the selective application of online platforms will not provide all the benefits of the digital transformation.

**Digital Marketing and Firm Performance**

Although with a relative importance score of 0.1587, Digital Marketing forms the third-highest predictor of firm performance with an explanatory power of approximately 20 percent lower than that of E-Commerce (0.1958), nonetheless, the predictor exhibits a significantly high effect compared to Cyber Security (0.0100) and Data Analytics (0.0032). Such a middle ground implies that although customer acquisition and customer engagement via the medium of digital marketing is a very important part of digital transformation, its efficiency is rather limited than other technological investments which are more basic to the core. The relative importance score shows that Digital Marketing has a significant contribution to performance prediction, although, it is mostly accompanied by a more critical predictive value such as Hardware/Software Investment and E-Commerce.

This quantification of the importance of Digital Marketing shows how it is a growth driver as well as brand differentiator as online market places continue to get busier. It is said that the 0.1587 score can be attributed to numerous value-creating phenomena: a rise in customer targeting with the

help of data-based campaigns, conversion rates with optimized digital touchpoints, and an increase in marketing ROI on measurable digital channels. Yet, the piece of information that it is not as significant as the infrastructure and e-commerce capabilities should be viewed as an indication that even the most advanced digital marketing instruments can bring little returns without proper technological outputs and transactional platforms. This follows the hypothesis of the digital marketing as a force multiplier that builds upon but never performs fundamentally the same digital infrastructure.

The ranking of the importance of Digital Marketing in the hierarchy has significant ramifications as far as the allocation of resources in digital transformation is concerned. Although the findings affirm its role as an effective performance lever, the results propose that before proceeding with a significant investment in advanced marketing technology, organizations must focus on developing basic and transactional digital capabilities. It also shows variation depending on the industry and those that focus more on the customer side might gain more relativity in the benefits of digital marketing than operationally-oriented sectors. Practitioners may want to note that there should be balanced digital portfolios in terms of both marketing excellence and strong digital platforms and ease of use of e-commerce capabilities.

### **Cyber Security and Firm Performance**

Having a relative importance value of 0.0100, Cyber Security is shown to be the least powerful predictor of the performance of firms, with significantly less explanatory power as digital core (Hardware/Software Investment: 1.0000) and customer-focused capabilities (E-Commerce: 0.1958; Digital Marketing: 0.1587). This low relative importance implies that, in contrast to a theoretical necessity in digital operations, cybersecurity measures have a minor direct impact on performance outcomes as compared to other bases of digital transformation. The score suggests an interpretation that cybersecurity acts more as an enabler than a factor directly drives the performance, and it is in line with its conceptualization as a “hygiene factor” in digital business models.

The quantification of importance is probably indicative of a number of dynamics. On the one hand, the investment in cybersecurity can provide only diminishing returns when basic protection levels have already been reached and further spending contributes to measured performance at a fraction. Second, the effects of cybersecurity can be primarily beneficial in preventing losses but not necessarily in increasing performances, and thus its positive effects are hard to quantify on measurable standards of performance. Third, cybersecurity value might not be specifically reflected in the measurement of other variables - e.g. strong security precautions might be a precondition to successful operation of e-commerce, so the effect of the same will be assigned to the more visible characteristics. This is consistent with the position in the latest research that the value of cybersecurity starts to show itself during crisis only and not during regular operations.

These results have significant implications in digital investment strategies. As the findings confirm the need of cybersecurity in the context of risk mitigation, they warn of overspending on security features to the detriment of digital capabilities that can enhance performance with a greater degree of immediateness. This could be due to a low relative significance, which can also include measurement constraints as typical financial indicators do not usually take into consideration prevented losses due to security penetrations. To the practitioners, this implies taking a more balanced approach where the cybersecurity is given the due attention, not an overwhelming one as compared to other digital transformation priorities. Future study ought to determine whether the value of cybersecurity is meaningfully different among industries with divergent risk exposures (e.g., finance and manufacturing).

### **Data Analytics and Firm Performance**

The least influential predictor in the data transformation components under analysis is Data Analytics with a relative importance score of 0.0032, thus it is found to have insignificant explanatory value when compared against other variables. Such negligible effect is quite impressive in the comparison with Hardware and Software Investment (1.0000) to which it has an insignificant 300 times less predictive power. The finding indicates that: in the existing motive of the modeling, the capability to make a decision based on data has minimal explanatory effect on variance in firm performance, although in the digital transformation literature, in its theoretical presentation, the capability to make data-driven decisions holds a prominent place.

There are some factors that can be attributed to this unreasonable outcome. To begin with, the advantages of analytics can be strongly dependent on other related organizational aspects (e.g. data literacy, managerial adoption) that are not included in the model. Second, analytics capabilities may add value essentially, as it drives other digital capability improvements (e.g. optimization of digital marketing or e-commerce), and therefore the contribution of analytics capabilities is included in those areas. Third, between the implementation of analytics and the new performance that can be measured, there can be big delays, especially when it comes to the more advanced solutions, such as predictive modeling. The findings are consistent with developing criticisms of the “data hype cycle”, which suggests that most organizations are trying to construe analytical potential capabilities to solid business value without institution of culture readiness cultural preparedness and data architecture.

Results of the research have significant qualifications for both in study and practice. Although they do not diminish the strategic potential of data analytics, they feature that its performance payoffs may be: (1) indirect and mediated by other digital capabilities, (2) dependent on unmeasured organizational variables, or (3) evident within subsets of functional areas or industry settings. To practitioners, this suggests explicit acceptance of the reality in analytics investments and the importance to develop the required pre-conditions - data governance systems and analytical talent pipelines, over which to expect expectation with potential performance improvements. Research needs to be done in future regarding the possibility that the key measurements may be stronger when examined within different timeframes (e.g., differentiating descriptive vs. predictive analytics) or when using a longer timeframe.

### Conclusion

In this article, the effects of five components of digitalization which included Hardware and Software Investment, Digital Marketing, E-Commerce, Data Analytics and Cyber Security on the performance of firms using complex predictive modeling and interpretation techniques were tested empirically. The analysis clearly showed that in terms of importance there was a clear hierarchy with the Hardware and Software Investment (relative importance = 1.0000) coming out as a major determinant, which is followed by the E-Commerce (0.1958), and Digital Marketing (0.1587) with the minor roles being played by Cyber Security (0.0100), and Data Analytics (0.0032). The validation of the model showed that it can present an extraordinarily great predictive accurate performance ( $R^2=0.989$ ), and further interpretation of the direction of contributions of all variables gained a lot through the use of Shapley value. Findings emphasize that underlying IT infrastructure, though, is the precondition to digital value creation, and that both customer-facing and analytical capabilities operate in nuanced ways, which may be indirect and not obvious. The findings can give theoretical explanation that favors a phased approach to digital transformation theory, as well as practical recommendations on where to focus strategic investments when pursuing an organizational digitalization process.

### References

- Bharadwaj, A. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly*, 24(1), 169–196.
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly*, 37(2), 471–482.
- Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing* (7th ed.). Pearson.
- Chong, A. Y. L., Li, B., & Liu, M. (2017). The impact of Industry 4.0 on supply chain performance: A study of the Chinese manufacturing sector. *Journal of Supply Chain Management*, 53(5), 16–29.
- Cui, J. (2025). AI-driven digital transformation and firm performance in Chinese industrial enterprises: Mediating role of green digital innovation and moderating effects of human-AI collaboration. arXiv preprint.
- Cui, Q., Yang, X., & Yang, J. (2020). E-commerce platforms and the reshaping of the retail industry. Springer.
- Davis, C., & Zhang, J. (2023). Integrating cybersecurity into operational processes: The moderating effect of organizational culture. *Information Systems Research*, 34(1), 67–82.
- Dedrick, J., Gurbaxani, V., & Kraemer, K. L. (2003). Information technology and business value: An empirical analysis of the role of IT in the productivity paradox. *Journal of Digital transformation and cybersecurity challenges for businesses*. (2022). *Sensors*, 23(15), 6666.

- Donthu, N., & Gustafsson, A. (2020). Effects of COVID-19 on business and research. *Journal of Business Research*, 117, 284–289.
- Dremel, C., Wulf, J., Herterich, M. M., Waizmann, J. C., & Brenner, W. (2017). How AUDI AG established big data analytics in its digital transformation. *MIS Quarterly Executive*, 16(2), 81–100.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2023). Digital transformation in manufacturing: Challenges and opportunities. *Business Process Management Journal*, 29(4), 873–891.
- Dutta, S., Seetharaman, P., & Mukherjee, S. (2020). Leveraging digital technologies for enhanced manufacturing performance: Evidence from Indian firms. *Asian Journal of Technology Innovation*, 28(2), 202–218.
- Fitzgerald, M., Kruschwitz, N., Bonnet, D., & Welch, M. (2014). Embracing digital technology: A new strategic imperative. *MIT Sloan Management Review*, 55(2), 1–12.
- Ghobakhloo, M., & Ching, N. T. (2019). Adoption of digital technologies in SMEs: A strategic framework. *Journal of Manufacturing Technology Management*, 30(8), 1236–1258.
- Laudon, K. C., & Traver, C. G. (2021). *E-commerce 2021: Business, technology, and society* (16th ed.). Pearson.
- Li, L., Su, F., Zhang, W., & Mao, J. Y. (2018). Digital transformation by SME entrepreneurs: A capability perspective. *Information Systems Journal*, 28(6), 1129–1157.
- Liu, Y., Zhou, F., & Xie, Q. (2023). Digital transformation and firm performance: A meta-analysis. *Technological Forecasting and Social Change*, 187, 122245.
- Lund, S., Madgavkar, A., & Manyika, J. (2021). Digital transformation and the post-pandemic recovery. *McKinsey Quarterly*, 2021(Spring).
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, 38, 97–102.
- Wade, M., & Hulland, J. (2004). The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly*, 28(1), 107–142.
- Xu, M., David, J. M., & Kim, S. H. (2018). The Fourth Industrial Revolution: Opportunities and challenges. *International Journal of Financial Research*, 9(2), 90–95.
- Yoo, Y., Boland, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, 23(5), 1398–1408.