Co-funded by the
Erasmus+ Programme
of the European Union

Building Skills 4.0 through University and Enterprise Collaboration

SHYFTE 4.0

WP1: WP Preparation

D1.1: Identify the skills requested by Industry4.0- Questionaries; workshop with Industry; relevant reports from EU commission

vs:2.0.0

Deliverable Lead and Editor: Matteo Savino and Marialuisa Menanno, UNS

Contributing Partners: ALL

Date: 04/02/2020

Dissemination: Confidential

Status: <~~Draft~~ ¦ Consortium Approved ~~¦ EU Approved~~>

This document evaluates the roles of core influencing factors on the benefits of Industry 4.0 in Asian industrial enterprises. The objectives are:

- Identify the determinants of enterprise budget relative to the introduction of Industry 4.0 innovations.
- Analyze how these determinants may evolve over time with respect to the adoption of Industry 4.0.
- Analyze the results of the proposed model using the Structure Equation Modelling

Project Ref: 598649-EPP-1-2018-1-FR-EPPKA2-CBHE-JP
Grant Agreement:
2018 - 3316 / 001 - 001

Co-funded by the
Erasmus+ Programme
of the European Union

Document Status

Deliverable Lead	Matteo Savino and Marialuisa Menanno, UNS
Internal Reviewer 1	A. Sekhari ULL
Internal Reviewer 2	J. Sarraipa UNL
Type	Deliverable
Work Package	WP1: WP Preparation on Industry 4.0
ID	D1.1:
Due Date	2020-04
Delivery Date	
Status	<~~Draft ¦~~ Consortium Approved ~~¦ EU Approved~~>

Status

This deliverable is subject to final acceptance by the EACEA.

Further Information

http://www.shyfte.eu/

Disclaimer

The views represented in this document only reflect the views of the authors and not the views of the European Union. The European Union is not liable for any use that may be made of the information contained in this document.

Furthermore, the information is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user of the information uses it at its sole risk and liability.

Project Partners:

Table of Contents

1. Executive Summary

The main objective of this document is to better understand the roles of core influencing factors on the benefits of Industry 4.0 in Asian industrial enterprises. The investigation of core management system application determinants, and the relationships between these determinants, is another purpose of this project.

This work was conducted and interpreted based on surveys from different areas and different industrial firms in Asia, based on Structural Equation Modeling (SEM) to analyze how these determinants can evolve over time and can affect employment of Industry 4.0 within the set of firms under investigation.

2. Principles and Fundamentals of Structural Equation Modelling (SEM)

2.1 Background of SEM

Structural equation modelling (SEM) is a powerful statistical technique for combining complex path models (PMs) with latent variable (LV) factors. It can specify confirmatory factor analysis (CFA) models, regression models, and complex path analysis (PA) models. Usually, SEM is used to build theoretical constructs with latent factors such as customer satisfaction and the influence of quality, which cannot be measured directly. SEM implies that there is a structure of the covariances among the observed variables. Moreover, the relationships among theoretical constructs are represented by path coefficients or regression among factors.

SEM development was first used for factor analysis to define a two-factor construct. Most of the achievements, diagnostic tests, surveys, and inventories used today were created using factor analytic techniques. The term “confirmatory factor analysis” is used to test the existence of these theoretical constructs.

PMs use correlation coefficients and regression analysis to analyse complex relationships among observed variables. The first applications of PA were designed for models of animal behaviours. In many respects, PA involves solving a set of simultaneous regression equations that theoretically establish the relationship among the observed variables. All of the hypothesized paths among those variables were shown to be statistically significant.

The final model type is SEM, which essentially combines PMs and confirmatory factor models (CFMs). SEM incorporate both latent and observed variables and became known as the linear structural relations model (LISREL) with the development of the first software program (LISREL) in 1973. Since then, many articles on SEM have been published. For example, Shumow and Lomax (2002) tested a theoretical model of parental efficacy for students.

Recently, the applications of SEM include factor analysis, measurement models, robustness, reliability and fit assessment, and interaction models, which has become a popular choice for analyzing multivariate methods.

To provide a basis for subsequent discussion, Shah and Goldstein (2006) presented two special cases that are frequently used in the operation management (OM) literature and a comprehensive detailing of the mathematical model specification. The research pointed to a distinction in the use of two terms that are often used interchangeably in OM: covariance structure modelling (CSM) and structural equation modelling (SEM). Moreover, SEM models are a subset of CSM models. But the current review is full of SEM models because other types of CSM models are rarely used in OM research.

SEM is a technique to specify, estimate, and evaluate the linear relationships among a set of observed variables in terms of a generally smaller number of unobserved variables. It builds a set of relations between one or more independent variables (IVs), which can be either continuous or discrete, and one or more dependent variables (DVs), which can be either continuous or discrete, to be examined.

SEM consists of observed variables (also called measured variables, MVs) and unobserved variables (also called latent variables, LVs), which can be independent (exogenous) or dependent (endogenous) in nature. LVs are hypothetical constructs that cannot be directly measured, which are typically represented by multiple MVs that serve as indicators of the underlying constructs in SEM. The SEM model is an a priori hypothesis about a pattern of linear relationships among a set of observed and unobserved variables. The objective in using SEM is to determine whether the a priori model is valid rather than to "find" a suitable model (Gefen et al., 2000).

SEM is also referred to as causal modelling, causal analysis, simultaneous equation modelling, analysis of covariance structures, PA, and CFA.

As shown in Figure 1. PA and CFA are two special cases of SEM that are regularly used in OM. PA models specify patterns of directional and non-directional relationships among MVs (Hair et al., 1998). Thus, PA provides for the testing of structural relationships among MVs when the MVs are of primary interest or when multiple indicators for LVs are not available. CFA requires that LVs and their associated MVs be specified before analysing the data. This is accomplished by restricting the MVs to load on specific LVs and by designating which LVs are allowed to correlate. A CFA model allows for directional influences between LVs and their MVs and (only) non-directional (correlational) relationships between LVs.

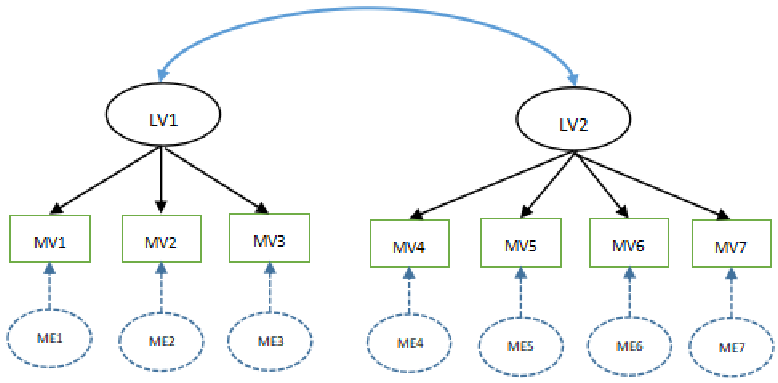

Figure 1. CFA models with two LVs and seven MVs.

A model of substance-use problems appears in Figure 1-2. Notice that the factors (often called LVs or constructs) are signified by circles. The observed variables (MVs) are signified by rectangles. These MVs could be items on a scale. Instead of simply combining the items into a scale by taking the sum or average of the items, creating a composite containing measurement error, the scale items are employed as indicators of a latent construct. Using these items as indicators of an LV rather than components of a scale allows for estimation and removal of the measurement error associated with the observed variables. This model is a type of SEM analysis called CFA. Often in later stages of research, after exploratory factor analysis (EFA), it is helpful to confirm the factor structure with data analysis using CFA techniques.

2.2 Key fit index of SEM

A traditional approach in SEM is to hypothesize a theoretical model, collect sample data, and test whether the model fits the data. In this section, we have discussed various fit indices to determine whether the theoretical model fits the data. When the theoretical model does not fit the data, we investigate modification indices for suggestions on how to modify the model with an improved fit.

The LV model of SEM shows the model fit and proposes the research hypotheses. In PA models, there are no LVs, and the hypotheses are represented by the paths among MVs. Like the measurement model fit, the sign, magnitude, and statistical significance of the structural path coefficients are examined in testing the hypotheses. Especially, the covariance fit is seen as more important than the variance fit. It is also important to distinguish these types of fits, because the model may fit well but be unable to explain the significant variation in endogenous variables.

As well as establishing the fit indices which meet all requirements, the report should include a variety of absolute and incremental fit indices for measurement, and moreover it should include a discussion of the interpretation of fit indices relative to the study design. In order to assess the model fit with numerous criteria, the model has been developed with different model-building assumptions in addition to the statistical approaches in multivariable procedures, such as the analysis of variance (ANOVA), multiple regression, discriminant analysis, multivariate ANOVA, and specified correlation analysis.

Many SEM model-fit indices are used to identify a correct model based on sample data and the observed variables which are assumed without error and statistical tests. A theoretical model in SEM is always a saturated model. The goal of SEM is to achieve a parsimonious model with a few substantive meaningful paths and a non-significant chi-square value close to the saturated model value of zero. Thus, it aims to indicate little difference between the sample covariance matrix and the reproduced implied covariance matrix. When the chi-square value is non-significant (close to 0), it means that the theoretical implied model fits the sample data.

Many of the model-fit criteria are computed based on knowledge of the saturated model, independence model, sample size, degrees of freedom, or chi-square values to formulate an index of model fit that ranges from 0 (no fit) to 1 (perfect fit). These various model fit indices are interpreted when determining an acceptable model fit. Some researchers have suggested that an SEM with a model-fit value of 0.90 or 0.95 or higher is acceptable (Baldwin, 1989). The various SEM programs report a variety of model-fit criteria based on output of LISREL. It is recommended that various model-fit criteria be used in combination to assess model fit, model comparison, and model parsimony as global fit measures.

In summary, we suggest that fit indices should not be regarded as a method of measurement of the usefulness of a model. The acceptable range of fit indices is shown in Table 1.

Table 1. Range of fit indices and interpretation

Model-fit criterion	Acceptable Level	Interpretation
Chi-square	Tabled χ^2 value	Compares $\chi 2$ with tabled value for df
Goodness-of-fit index (GFI)	0 (no fit) to 1 (perfect fit)	Good fit value close to .90 or .95
Adjusted GFI (AGFI)	0 (no fit) to 1 (perfect fit)	Good model fit with .90 or .95 for value adjusted df
Root-mean square residual (RMR)	Researcher defines level	Indicates the closeness of Σ to the S matric
Standardized RMR (SRMR)	< 0.05	Value less than .05 with a good model fit
Root-mean-square error of approximation (RMSEA)	0.05 to 0.08	Good fit value close to .90 or .95
Normed fit index (NFI)	0 (no fit) to 1 (perfect fit)	Good fit value close to .90 or .95
Non-normed fit index (NNFI)	0 (no fit) to 1 (perfect fit)	Good fit value close to .90 or .95

These criteria are used to judge the statistical significance and substantive meaning of a theoretical model. They concern the non-statistical significance of the chi-square test, the statistical significance of individual parameter estimates, and the magnitude and direction of the parameter estimates (which aims to investigate whether a positive or negative coefficient is influenced by the parameter estimate). Also, they reflect the “absolute” fit and the model's “incremental” fit. Absolutely, indicators of model fit include χ^2 and SRMR, among others. Incremental fit statistics include the comparative fit index (CFI) among others.

These model-fit statistics can be expressed in terms of the non-centrality parameter (NCP). The estimate of NCP using the maximum likelihood (ML) chi-square is $\chi^2 - df$. Some of the fit indices are computed given knowledge of the null model χ^2 (independence model, where the covariance terms are assumed to be zero in the model), null model df , hypothesized model χ^2 , hypothesized model df , number of observed variables in the model, number of free parameters, and sample size. The formulae for the goodness-of-fit index (GFI), normed fit index (NFI), relative fit index (RFI), incremental fit index (IFI), comparative fit index (CFI), and RMSEA using these values are as follows:

- *Chi-Square χ^2*

A significant χ^2 value relative to the degrees of freedom indicates that the observed and implied variance covariance matrices differ. Statistical significance indicates the probability that this difference is due to sampling variation. A non-significant χ^2 value indicates that the two matrices are similar, which implies that the SEM model is significantly based on sample covariance relationships in the matrix.

The chi-square test of model fit can lead to erroneous conclusions regarding analysis outcomes. The χ^2 model-fit criterion is sensitive to sample size because as the sample size increases (generally above 200), the χ^2 statistic has a tendency to indicate a significant probability level. In contrast, as the sample size decreases (generally below 100), the χ^2 statistic indicates non-significant probability levels.

Three estimation methods are commonly used to calculate χ^2 in LV models (Loehlin, 1987). Each approach estimates a best-fitting solution and evaluates the model fit. Facts about χ^2 are detailed as below:

1) It increases as a function of df. If the model fits extremely well, or a sample has a size of 2000, the χ^2 would be tent to 2000 approximately.

2) χ^2 ranges from 0 to a high value. It is 0 when the saturated model is fit (all possible paths are in the model to be estimated). It is at its highest on any data set for the model of independence (no paths are entered into the model).

3) χ^2 penalizes models with a large number of variables (it is large when there are many variables).

4) χ^2 decreases as parameters are added to the model (much like an R2 would increase as predictors are added). However, adding parameters means that the model is becoming more complex and less parsimonious.

5) χ2 can be used to compare the fits of nested competing models.

where model A is a restricted version of B. The result is a distributed χ^2 with degrees of freedom equal to $df_A - df_B$. If model A is nested in model B, B estimates more parameters, whereas in model A, more parameters are fixed (usually to 0) and not estimated. If two models are not nested, they will use descriptive goodness-of-fit measures, such as an adjusted goodness-of-fit index (AGFI).

Goodness-of-fit index and adjusted goodness-of-fit index

The GFI is based on the ratio of the sum of the squared differences between the observed and reproduced matrices to the observed variances, thus allowing for scale. The GFI measures the amount of variance and covariance in S that is predicted by the reproduced matrix Σ.

The GFI index can be computed for ML estimates (Bollen, 1989). For our modified model, the formula is expressed as:

$$GFI = 1 - [{\chi_{\mathrm{model}}}^2 / {\chi_{null}}^2]$$

(Note: χ^2_{null} is the chi-square for the independence model with degrees of freedom.)

The AGFI is adjusted for the degrees of freedom of a model relative to the number of variables. The AGFI is computed as

$$AGFI = 1 - [(k/df)(1 - GFI)]$$

where k is the number of unique distinct values in S, which is $p(p+1)/2$. df is the number of degrees of freedom.

The GFI and AGFI indices can be used to compare the fit of two different models with the same data or to compare the fit of a single model using different data, such as separate data sets for males and females.

- *Root-mean-square residual (RMR) index*

RMR means the root-mean-square residual. The differences between the data in S and the model in Σˆ are called residuals. The average of these residuals is calculated on how data is far off the model. The square root of that value is considered to calculate the index of the “standard deviation” scale rather than a “variance” scale. The matrices S and Σ ˆ are typically covariance matrices; moreover the index is more easily interpreted if it is standardized (as if it were computed on a correlation matrix where the variances were equal to 1.0), so that it ranges from 0.0 to 1.0. The equations for the RMR and SRMR (the standardized root-mean-square residual) are shown as follows (Browne et al., 2002):

$$RMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \sigma_{ij})^2}{k(k+1)/2}}$$

where $k = p + q$

$$SRMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}[(s_{ij} - \sigma_{ij})/(s_{ii}s_{jj})]^2}{k(k+1)/2}}$$

χ^2, RMR and SRMR are badness-of-fit indices—higher values indicate worse fits. If the model predicts the data fairly closely, then the residuals should be close to 0, making the numerator of RMR obviously 0 and the numerator of SRMR close to 0.

The RMR index also uses the square root of the mean-squared differences between matrix elements in S and Σ. It is calculated as:

$$RMR = [(1/k)\sum_{ij}(s_{ij} - \sigma_{ij})^2]^{1/2}$$

The SRMR has an acceptable level when it is less than 0.05. For this index, 1) Hu and Bentler (1995) suggest that an SRMR “close to .09” represents a reasonable fit; 2) through simulation testing, the SRMR has been characterized as more sensitive to model misspecification than to sample size. Thus, if the SRMR is not as low as desired, the inflation is a fairly clear indicator that something is wrong with the (measurement or structural) model.

- *Root-mean-square error of approximation (RMSEA)*

MacCallum et al. (1996) provided a different approach for testing model fit using the RMSEA. It is an index like the SRMR but is computed differently (Steiger, 2000).

$$RMSEA = \sqrt{(X^2 - df)/df(N-1)}$$

Alternatively it can be calculated as:

$$RMSEA = \sqrt{NCP\ /\ df\,(N-1)}$$

In simulation studies, RMSEA over-rejects true models for “small” N (N = 250), and the fit tends to worsen as the number of variables in the model increases (Kenny and McCoach, 2003). Thus SRMR is preferred. The approach also emphasizes confidence intervals around RMSEA, rather than a single point estimate (RMSEA ≤ .05).

- *Normed fit index (NFI) and comparative fit index (CFI)*

The CFI was developed more than 30 years ago. The problem of large χ^2 and a non-informative state of rejecting the null hypothesis led researchers to develop other model evaluation criteria. In particular, Bentler and Bonett (1980) discovered that an index should compare a model's fit not against a straw-model (null model) but against an idealized (yet still simple) model. Thus, these statistics became known as model comparison or incremental fit indices (Bentler 1990).

The NFI is defined as follows:

$$NFI = \sqrt{(\chi_{null}{}^2 - \chi_{\mathrm{model}}{}^2)/\chi_{null}{}^2}$$

It ranges from 0.0 to 1.0. The χ^2-model is the fit of the model and the χ^2 null model is the fit of the model of independence which estimates variances but no covariances (there are no paths in the model between any constructs and all the variables are thought to be independent).

NFI is influenced by sample size and underestimates fit in small samples. It is difficult to compare across data sets (Ding et al., 1995; Arbuckle et al., 1996). Thus a new index was created to correct these shortcomings. The CFI generated ranges from 0.0 to 1.0. It defines:

$$CFI = 1 - \left\{ [Max((\chi_{\mathrm{model}}{}^2 - df_{\mathrm{model}}),0)]/[Max((\chi_{null}{}^2 - df_{null}),0)] \right\}$$

The comparison (by subtraction) of a model's χ^2 and its df is an adjustment for model parsimony. Models tend to fit worse (χ^2 is larger) when few parameters are estimated (when there are many df). Yet if a model fits well (χ^2 is small), there is a penalty if that fit is achieved through a complex model (one with many parameters, using many df). Then, the comparison (by ratio) of the focal model to the null model reflects the extent to independence of current data. Instead, if there were nothing going on within the data and the independence model were true, the values of χ^2 (for the null model) would be similar. If the df were similar, the entire ratio would be approximately 1.0, and hence CFI = 1 – 1, which is 0.0. Thus, a CFI gets larger as the model and data become more interesting, moving away from a simplistic model of independence.

The CFI has been said to be somewhat forgiving in exploratory modelling (Rigdon, 1996). Overall, Hu and Bentler (1998) have demonstrated strong performance (power and robustness) of the CFI.

Bentler (1990) subsequently developed a coefficient of comparative fit within the context of specifying a population parameter and distribution, such as a population comparative fit index, to overcome the deficiencies in NFI for nested models. The rationale for assessment of comparative fit in the nested-model approach involves a series of models that range from least restrictive (Mi) to saturated (Ms). Corresponding to this sequence of nested models is a sequence of model-fit statistics with associated degrees of freedom. The CFI measures the improvement in non-centrality from model Mi to Mk (the theoretical model) and uses the non-central χ^2 distribution with the non-centrality parameter lk to define CFI as:

$$CFI = (li - lk)/li$$

McDonald and Marsh (1990) further explored the non-centrality and model-fit issue by examining nine fit indices as functions of non-centrality and sample size. Model fit determines the degree to which the sample variance–covariance data fit the SEM. Commonly used model-fit criteria are chi-square (χ2), GFI, AGFI, and RMR (Jöreskog and Sörbom, 1993). These criteria are based on differences between the observed (original, S) and model-implied (reproduced, Σ) variance–covariance matrices.

Mulaik et al. (1989) evaluated the χ^2, NFI, GFI, and AGFI indices. They concluded that these indices fail to assess parsimony and are insensitive to misspecification of structural relationships. Moreover, it has been suggested that a good fit index is independent of sample size, accurately reflects differences in fit, imposes a penalty for inclusion of additional parameters, and supports the choice of the true model when it is known (McDonald and Marsh, 1990). No model-fit criteria can actually meet all of these criteria.

Following initial description, there has been much controversy and discussion on their subjective interpretation and appropriateness under specific modelling conditions (Marsh, Balla, and Hau (1996). Kenny and McCoach (2003) indicated that RMSEA improves as more variables are added to a model, whereas CFI decline in correctly specified models as more variables are added.

When making ideas to report model-fit indices, considering the fit indices were suitable for model fit, model parsimony, model comparison or not firstly. For example, the CFI should be reported for comparing models.

To avoid the risk of oversimplification, χ^2, RMSEA, and SRMR would be reported for all types of models. Overall, more than one model-fit index should be reported. If a majority of the fit indices indicate an acceptable model, then the theoretical model will be supported by the data.

2.3 Path model and confirmatory factor model of SEM

The core model analysis of SEM path model (PM) and confirmatory factor model (CFM) by using multiple regression. Multiple regression is a general linear modelling approach to the analysis of data to solve the gap between correlation and ANOVA in answering research hypotheses. Many researchers investigate the relationships between multiple regression and variance analysis (Lomax, 1982).

Path analysis (PA)

PA uses models involving multiple observed variables such as independences and DVs of equations. Thus, it requires multiple regression equations using observed variables.

PA is a method of studying the direct and indirect effects of variables. It is not only for discovering causes but also tests theoretical relationships, which historically has been termed causal modelling. A specified PM establishes causal relationships among two variables when:

- temporal ordering of variables exists;
- covariance or correlation is present among variables;
- other causes are controlled for;
- a variable X is manipulated, which causes a change in Y.

Pearl (2009) has renewed a discussion of causation in the behavioral sciences with model examples and the rationale for causation which can be expressed in mathematical expressions ready for computer analysis and fits into the testing of theoretical PMs.

Once a particular PM has been specified, the next concern is whether the model is identified. In SEM, it is crucial that the researcher resolve the identification problem prior to the estimation of parameters.

Model specification is necessary in examining multiple variable relationships in PMs, just as in the case of multiple regression. Many different relationships among a set of variables can be hypothesized with many different parameters being estimated. In a simple three-variable model, for example, many possible PMs can be postulated on the basis of different hypothesized relationships among the three variables.

This is known as model specification and shows the important role that theory and previous research play in justifying a hypothesized model. PA does not provide a way to specify the model but rather estimates the effects among the variables once the model has been specified a priori by the researcher on the basis of theoretical considerations. For this reason, model specification is a critical part of SEM modelling.

In multiple regression, the DV is regressed in a single analysis on all of the IVs. In PA, one or more multiple regression analyses are performed depending on the variable relationships specified in the PM. Path coefficients are therefore computed only on the basis of the particular set of IVs that lead to the DV under consideration.

Confirmatory factor models (CFM)

CFM was developed for involving factors or LVs modelling. In this part, a major limitation of models involving only observed variables is that measurement error is not taken into account. The use of observed variables in statistics assumes that all of the MVs are perfectly valid and reliable, which is unlikely in many applications. For example, people's educational level is not a perfect measure of a socioeconomic status factor and amount of exercise per week is not a perfect measure of a fitness factor.

The validity and reliability issues in measurement have traditionally been handled by first examining the validity and reliability of scores on instruments used in a particular context. Given an acceptable level of score validity and reliability, the scores are then used in a statistical analysis. However, the traditional statistical analysis of these scores—for example, in multiple regression and PA—does not adjust for measurement error. The impact of measurement error has been investigated and found to have serious consequences—for example, biased parameter estimates. SEM software that accounts for the measurement error of variables was therefore developed—that is, factor analysis—which creates LVs used in SEM.

Factor analysis attempts to determine which sets of observed variables share common variance–covariance characteristics that define theoretical constructs or LVs. Factor analysis presumes that some factors that are smaller in number than the number of observed variables are responsible for the shared variance–covariance among the observed variables. In practice, one collects data on observed variables and uses factor-analytic techniques to either confirm that a particular subset of observed variables defines each construct or factor or to explore which observed variables relate to factors. In exploratory factor model approaches, we seek to find a model that fits the data, so we specify different alternative models, hoping to ultimately find a model that fits the data and has theoretical support. This is the primary rationale for exploratory factor analysis (EFA). In CFM approaches, we seek to statistically test the significance of a hypothesized factor model—that is, whether the sample data confirm that model. Additional samples of data that fit the model further confirm the validity of the hypothesized model. This is the primary rationale for CFA.

In CFA, the researcher specifies a certain number of factors, which factors are correlated, and which observed variables measure each factor. In EFA, the researcher explores how many factors there are, whether the factors are correlated, and which observed variables appear to best measure each factor. In CFA, the researcher has an a priori specified theoretical model; in EFA, the researcher does not have such a model.

2.4 Advantages of SEM

(i) Traditional statistical methods have only a very limited number of variables, and now the research needs to use more observation variables.

ii) Structural equations can be modeled and tested for complex objects.

iii) SEM is tolerant to measurement errors, multi-level SEM modelling and multi-level data collection can be used, and SEM provides analysis of complex phenomena to solve the problem.

iv) SEM software been developed and is getting easier and faster to use.

SEM is a correlation research method. Basically, researchers should know their data characteristics. Data screening and preparation is a very important first step in SEM. The next paragraph discusses detailed issues related to the use of correlation and variance, which play important roles in SEM models.

The correlation procedure was developed for factor analysis techniques at the beginning. The correlation, regression, and factor analysis techniques have for many decades formed the basis for generating tests and defining constructs. Then, the Pearson correlation coefficient provides the basis for point estimation (test of significance), explanation (variance accounted for in a DV by an IV), prediction (linear regression from an IV to a DV), reliability estimates (test-retest), and validity (factorial, predictive, concurrent).

Given the important role that correlation plays in SEM, the key factors that affect the establishment of relationships among multi-variable data points are the level of measurement, restriction of range of data values (variability, skewness, kurtosis), missing data, non-linearity, outliers, correction for attenuation, and issues related to sampling variation, confidence intervals, effect size, significance, sample size, and power.

Initially, SEM required variables measured at the interval or ratio level of measurement, so the Pearson product-moment correlation coefficient was used in analysis of regression, path, factor, and SEM. The Pearson correlation coefficient indicates the degree of linear relationship between two variables. It is possible that two variables can indicate no correlation if they have a curvilinear relationship. SEM software programs estimate coefficients based on the user-specified theoretical model or implied model, but also must work with the saturated and independence models.

Ding et al., (1995) located numerous studies that were in agreement that the minimum satisfactory sample size is 100 to 150 subjects when conducting SEM. Hu, Bentler, and Kano (1992) indicated that in some cases 5,000 is insufficient. Costello and Osborne (2005) demonstrated in their Monte-Carlo study that it is recommended that the best practice is to use 20 subjects per variable in factor analysis. But with examination experience of this study in addition to published SEM research, we agreed many articles' finding that 250 to 500 subjects, or the greater the sample size, it is more likely that the subjects can validate the model. For example, Bentler and Chou (1987) suggested that a ratio as low as five subjects per variable would be sufficient for normal and elliptical distributions when the LVs have multiple indicators and that a ratio of at least 10 subjects per variable would be sufficient.

As previously discussed, the Pearson correlation coefficient is limited by the range of score values and the assumption of linearity, among other things. Even if the assumptions and limitations of using the Pearson correlation coefficient meet the requirements, a cause-and-effect relationship still has not been established. Thus in such a case, association rather than causation can be inferred.

3. Shyfte 4.0 Methodology

The purpose of this document "Shyfte 4.0", is to the evaluate the actual technological and organizational benefits related to the adoption of Industry 4.0 technologies, based on Structural Equation Modeling (SEM) to analyze how these determinants can evolve over time and can affect employment of Industry 4.0 within the set of firms under investigation.

Structural equation models are used to evaluate unobservable "latent" constructs. They are often based on a measurement model that defines latent variables using one or more observed variables and a structural model that considers the relationships between latent variables (Azevedo and Ferreira, 2007). SEM can examine different dependency relationships simultaneously (Savino and Shafiq, 2018; Sadia et al., 2018; Srivastava and Dubey, 2014).

To explore the potential benefit of Industry 4.0, we have designed the quantitative research study of Figure 1. According to this model, the potential benefit of Industry 4.0 can be defined as the individual capacity perceived in the implementation of Industry 4.0. The design of our study contains six hypotheses shown in Fig. 2. As shown in the literature review, many studies concern the potential benefits of Industry 4.0 relating to the use of new technologies (Schimidt et. al., 2015). Thus, we considered the main factors and investigated them to assess the potential benefits found in the implementation of Industry 4.0, such as Implementation level of Industry 4.0, Importance of Industry 4.0, Technologies and software systems, Skill requirement, Investment Sector and Artificial Intelligence.

Basl (2017), in her survey on the availability of companies in implementing the principles of Industry 4.0 concludes that greater implementation of the principles of Industry 4.0 in companies is still hampered by the high costs associated with the application of Industry 4.0 solutions.

The determinants are also chosen through the expert evaluations conducted through a questionnaire. The experts who answered the questionnaire deal with the digitalization of production processes based, for example, on the use of the Internet of Things devices and solutions that autonomously communicate with each other along the entire value chain.

According to the findings of the literature (Gerbert et al., 2015; Lu, 2017) and as regards to a preliminary investigation conducted in a sample of firms, the model is provided of the following hypotheses:

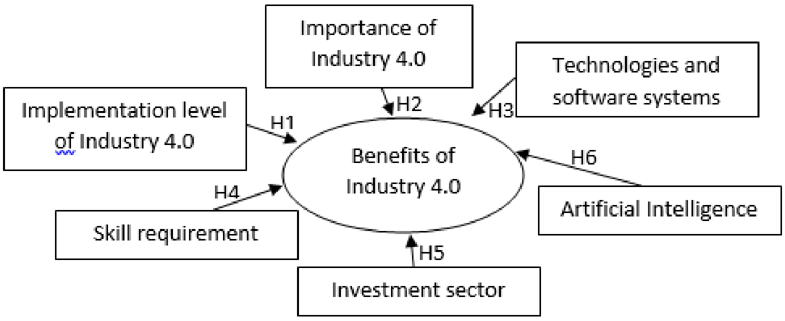

Figure 2: Hypotheses and test model

- H1: Implementation level of Industry 4.0 has a positive impact on benefits.
- H2: The importance of Industry 4.0 has a positive impact on benefits.
- H3: Technologies and software systems have a positive impact on benefits of Industry 4.0.
- H4: Skill requirement has a positive impact on benefits of Industry 4.0.
- H5: Investment sector has a negative impact on benefits of Industry 4.0.
- H6: Artificial Intelligence has a negative impact on benefits of Industry 4.0.

Data collection and demographic distribution of the samples:

Data collection involved a sample of 300 companies based in China, Malaysia and Thailand. The survey was conducted in the second half of 2019. The comparative model was set through a questionnaire (Table 2) composed of 20 questions grouped into seven sections. The first six groups of questions are relative one of the six aforementioned factors. The responses are organized according to a five-point Likert scale, with a score ranging from 1 (‘strongly disagree’) to 5 (‘strongly agree’) (Thirupathi and Vinodh, 2016).

Table 2. The questionnaire of the survey.

Factors	Questions
Implementation Level	Q1. The vision of the digital transformation
	Q2. I4.0 technologies
	Q3. Willingness to adopt I4.0 technologies
	Q4. Investments in the implementation of I4.0
Importance of Industry 4.0	Q5. Importance of I4.0 technologies
	Q6. Importance of Skills among employees
	Q7. Investments and threatening for SMEs

Investment sector	Q8. Investments for the future in the realization of I4.0
Technologies and software systems	Q9. Technologies used in wireless networks
	Q10. Type of data and technology
	Q11. Programing Language
	Q12. Type of data analytics
Skill requirement	Q13. Skills development
Artificial Intelligence	Q14. Type of AI algorithm
	Q15. Type of technology
	Q16. Programming Language
Benefits of Industry 4.0	Q17. Capital
	Q18. Productivity
	Q19. Market share
	Q20. Quality of your products/services

3.1 Shyfte 4.0 Data Collection and Analysis

The questionnaire was sent to 95 companies based in China, 93 companies based in Malaysia and 100 companies based in Thailand. The person interviewed of each company was the production manager, Chief Executive Officer, Chief Information Officer, Director and Engineer depending on the availability of the firm surveyed, thus getting one response from each company. We have required those kind of professionals because of their specific competencies on quality and production efficiency required for this investigation.

This portion of the research returned 87 responses from China companies, 86 from Malaysia and 92 responses from Thailand companies after excluding invalid responses. Therefore, 265 usable responses were obtained at a response rate of 92%.

Considering that response rates of large-scale surveys are often about 5%–10% (Alreck and Settle, 1995), the response rate can be considered as acceptable. The potential bias due to the non-respondent was tested through the difference between the two types of respondents. The main assumption is that the second type of respondents may act in the same way as non-respondents (Wu et al., 2008). All the paired-sample t-tests conducted for all the variables showed a high difference in the mean values. Thus, the nonresponse bias is not significant for the study. The respondents were also clustered with respect to the number of employees (Table 3), the findings show that most of the companies consist of more than 100 employees.

Table 3. Survey clustering with respect to the number of employees.

Cluster	Number of respondents	Rate (%)
>=100	162	61
From 20 to 99	66	25
<20	37	14

The validation of the construct was conducted using confirmatory factor analysis (CFA) with EQS software Rev 6.3. The main univariate results are reported in Table 4, with mean, standard deviation and skewness.

Table 4. Univariate results.

	Resource	Question	Mean	Standard deviation	Skewness
1	Implementation Level	Q1	2.735	1.068	0.150
		Q2	2.406	0.923	0.109
		Q3	2.396	0.899	0.103
		Q4	2.421	0.949	0.136
2	Importance of Industry 4.0	Q5	2.476	0.899	0.034
		Q6	2.404	0.911	0.132
		Q7	2.437	0.938	0.184
3	Investment sector	Q8	2.433	0.944	0.072
4	Technologies and software systems	Q9	2.439	0.973	0.189
		Q10	2.423	0.927	0.132
		Q11	2.525	0.960	0.174
		Q12	2.374	0.910	0.147
5	Skill requirement	Q13	2.599	0.790	0.186
6	Artificial Intelligence	Q14	2.429	0.893	0.062
		Q15	2.417	0.894	0.096

	Resource	Question			
			Mean	Standard deviation	Skewness
		Q16	2.569	1.070	0.095
7	Effects of Industry 4.0	Q17	2.340	0.930	0.171
		Q18	2.554	0.988	0.060
		Q19	2.923	1.081	0.071
		Q20	2.450	1.144	0.121

Within the validity check of the model, our results were consistent with the findings of Sila and Ebrahimpour (2005) that were further discussed by Bagozzi (2010). According to these authors, the empirical evidence in CFA is generally appraised using indices such as the Bentler–Bonett Normed Fit Index (NFI), comparative fit index (CFI), χ2 test, Standardized RMR, Root Mean-Square Error of Approximation (RMSEA) and Non Normed Fit index (NNFI). The RMSEA demonstrates an adequate model fit if it is less than 0.08 (Wu et al., 2008) and a better fit if less than 0.06 (Bagozzi, 2010). CFI value is accepted if it is greater than 0.9, although a better fit is for values above 0.95 (Iacobucci, 2010). Table 5 reports the values obtained, along with the relative threshold values. The overall CFA showed a fairly good fit level, as indicated by CFI = 0.967, NFI = 0.946, NNFI = 0.964, IFI = 0.976, GFI = 0.908, RMR = 0.060, RMSEA = 0.0076.

Table 5. CFA analysis

INDICES	RMR	RMSEA	CFI	GFI	IFI	NFI	NNFI
Fit value	<0.05	<0.08	>0.9	>0.9	>0.9	Close to 1	Close to 1
Shyfte Results	0.040	0.0076	0.967	0.908	0.967	0.946	0.964

All fit indices are accepted and significant at χ2 = 1012.667, thus confirming the convergent validity. The discriminant validity was assessed by two-factor CFA (Wu et al., 2008, Utriainen et al., 2018), generating sub-models composed of all pairs of constructs. This portion of the study generated 28 models. In all these models, the χ2 of the unconstrained model was lower than the χ2 of the constrained models, thus confirming the discriminant validity. The Cronbach-α (C-α) coefficient is also used to measure the reliability of a construct. C-α≥ 0.7 can be deemed of good scale reliability (Nunnally, 1978, Wu et al., 2008). In this study, for all the constructs the C-α ranged between 0.87 and the maximum value of 0.95, also indicating the reliability of the constructs. The survey was conducted with a single respondent from each company and at the same time. Hence, the study was also tested using common method bias (Podsakoff et al., 2003). Harman's single factor test was applied in CFA by loading all the measurements on one latent variable. The analysis generated a value of χ2 = 3135.620 with a degree of freedom (df) equal to 349. Then, the value of χ2 was compared with the value obtained using the full model (χ2 = 1012.667 with df = 322). This result allowed rejection of the potential hypothesis regarding a potential single factor accounting for the most part of variance. The constructs showed fair measurement properties and were able be used to test the hypotheses.

Fig. 3 reports the multivariate results of the model, with the relative path coefficients. As all coefficient factors are positive, however the basic resources have not all the same weight, thus that they do not impact towards the hypotheses to the same extent.

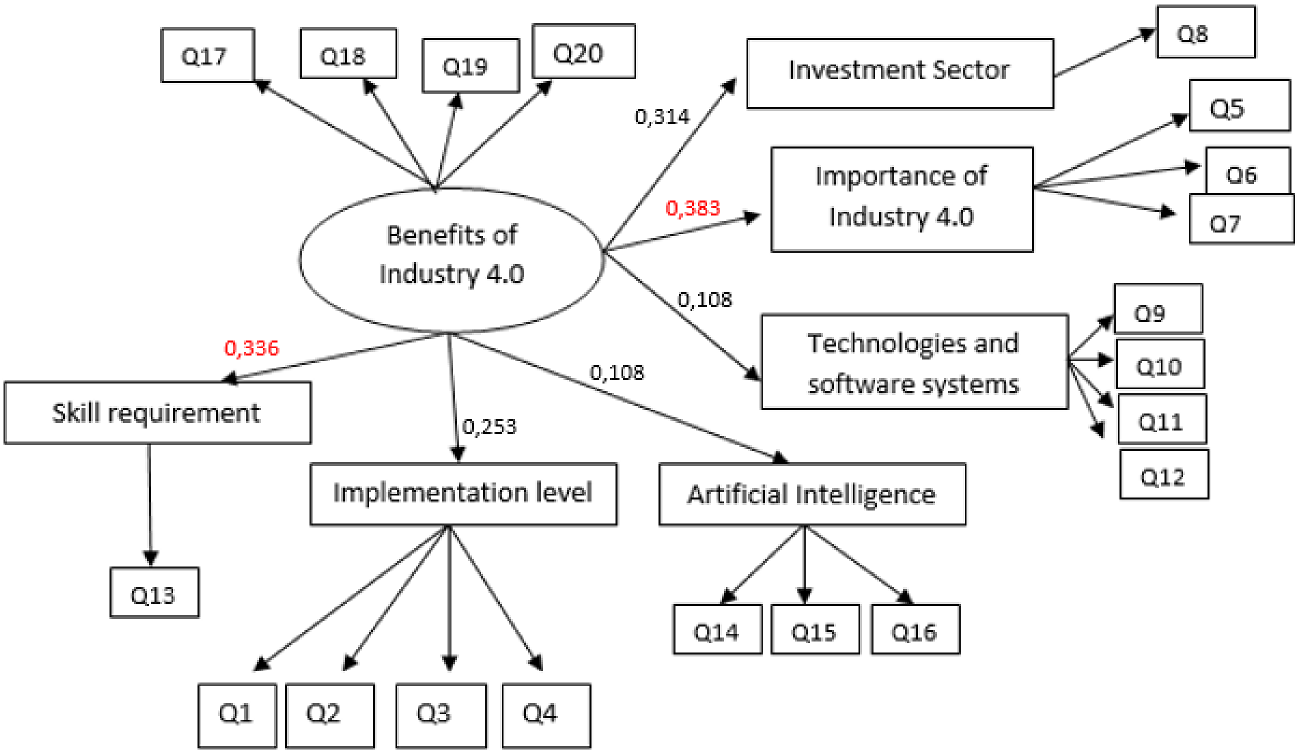

Figure 3. Multivariate analysis

Analysis of the model shows that:

- Hypothesis 1 (H1) suggests that implementation level of Industry 4.0 has a positive impact on benefits of Industry 4.0 implementation by companies. The empirical results indicate a significant and positive relationship between implementation level and benefits of Industry 4.0. Hence, H1 is supported.
- Hypothesis 2 (H2) states that the ~~of~~ Industry 4.0 has a positive impact on benefits of Industry 4.0 implementation by companies. The empirical results reveal a highly significant and positive relationship between importance and benefits of Industry 4.0. Thus, H2 is strongly supported.

- Hypothesis 3 (H3) suggests that technologies and software systems have a positive impact on benefits of Industry 4.0 implementation by companies. Again, the empirical results show a significant and positive relationship between technologies and software systems and benefits of Industry 4.0. Consequently, H3 is supported.
- Hypothesis 4 (H4) suggests that skill requirement has a positive impact on benefits of Industry 4.0 implementation by companies. The empirical results indicate a highly significant and positive relationship between skill requirement and benefits of Industry 4.0 implementation. Hence, H4 is strongly supported.
- Hypothesis 5 (H5) suggests that investment sector has a negative impact on benefits of Industry 4.0 implementation by companies. The empirical results show a significant and positive benefits. Thus, H5 is opposite the prediction.
- Hypothesis 6 (H6) suggests that Artificial Intelligence has a negative impact on benefits of Industry 4.0 implementation by companies. The empirical results show a significant and positive benefits. Thus, H6 is opposite the prediction.

The path coefficients are summarized in Table 6.

Table 6. Hypotheses test results

Hypotheses	Path coefficient	Result
H1: Implementation level of Industry 4.0 has a positive impact on benefits.	0.253	Support
H2: The importance of Industry 4.0 has a positive impact on benefits.	0.383	Strongly Support
H3: Technologies and software systems have a positive impact on benefits of Industry 4.0.	0.108	Support
H4: Skill requirement has a positive impact on benefits of Industry 4.0.	0.336	Strongly Support
H5: Investment sector has a negative impact on benefits of Industry 4.0.	0.314	Reject
H6: Artificial Intelligence has a negative impact on benefits of Industry 4.0.	0.108	Reject

3.2 Shyfte 4.0 Contingency effects for different Analysis

We have analyzed other contingency effects by considering i) the different number of employees, and ii) location of the company.

Regarding the first analysis, the first group contained 92 observations while the second and third contained 86 observations. The analysis has been conducted for all groups to analyse the trend of the path coefficients referring to each resource.

The resulting fit statistics of the modified model were satisfactory with NFI = 0.94, CFI = 0.95, RMSEA = 0.05; NFI = 0.91, CFI = 0.93, RMSEA = 0.57, for the second group and NFI = 0.94, CFI = 0.94, RMSEA = 0.59, for the third group.

The core resources and the relative values of the path coefficients are reported in Table 7, in which higher values are highlighted in green and the hypotheses that have been rejected are highlighted in red.

Table 7. Core resources and path coefficients

Hypotheses	Path coefficient			Result
	Number of Employees <20	Number of Employees from 20 to 99	Number of Employees >99	
H1: Implementation level of Industry 4.0 has a positive impact on benefits.	0.345	0.256	0.137	Support
H2: The importance of Industry 4.0 has a positive impact on benefits.	0.484	0.332	0.294	Strongly Support
H3: Technologies and software systems have a positive impact on benefits of Industry 4.0.	0.040	0.166	0.120	Support
H4: Skill requirement has a positive impact on benefits of Industry 4.0.	0.432	0.385	0.230	Strongly Support
H5: Investment sector has a negative impact on benefits of Industry 4.0.	0.378	0.259	0.269	Reject
H6: Artificial Intelligence has a negative impact on benefits of Industry 4.0.	0.011	0.026	0.149	Reject

The Second analysis has been made considering the number of responses obtained from different countries.

Table 8. Core resources and path coefficients

Hypotheses	Path coefficient			Result
	MALAYSIA	THAILAND	CHINA	
H1: Implementation level of Industry 4.0 has a positive impact on benefits.	0.292	0.570	0.240	Support
H2: The importance of Industry 4.0 has a positive impact on benefits.	0.489	0.142	0.275	Support
H3: Technologies and software systems have a positive impact on benefits of Industry 4.0.	0.179	0.027	0.088	Support
H4: Skill requirement has a positive impact on benefits of Industry 4.0.	0.461	0.126	0.377	Support
H5: Investment sector has a negative impact on benefits of Industry 4.0.	0.364	0.142	0.265	Reject
H6: Artificial Intelligence has a negative impact on benefits of Industry 4.0.	0.100	0.064	0.143	Reject

3. 3 Process and analysis of statistics data

The elaboration and analysis of the answers of the interviewees are also carried out considering the percentage values for each question of the questionnaire.
By evaluating the questionnaire previously considered and adding information about the company, we summarize the results below.

Company Capital

1 = (< €0.5 Million)	14%
2 = (> €5 < €15 Million)	28%
3 = (> 15Million)	59%

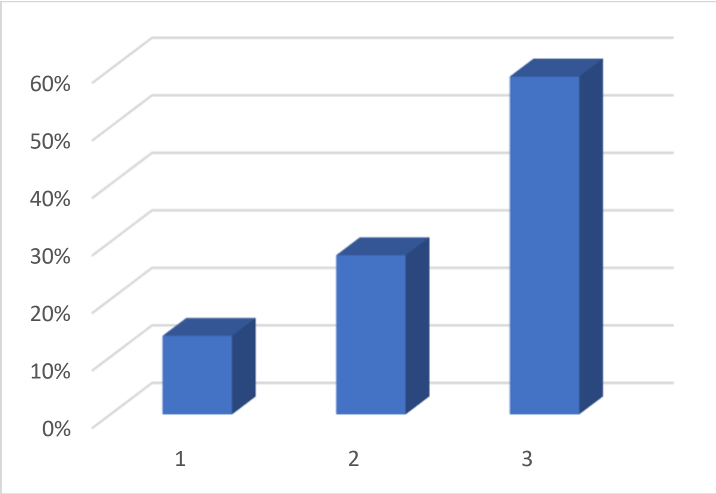

Figure 4. Company capital distribution

The size of company (number of employees)

1 = (Up to 20 employees)	14%
2 = (From 20 to 99 employees)	25%
3 = (More than 100 employees)	61%

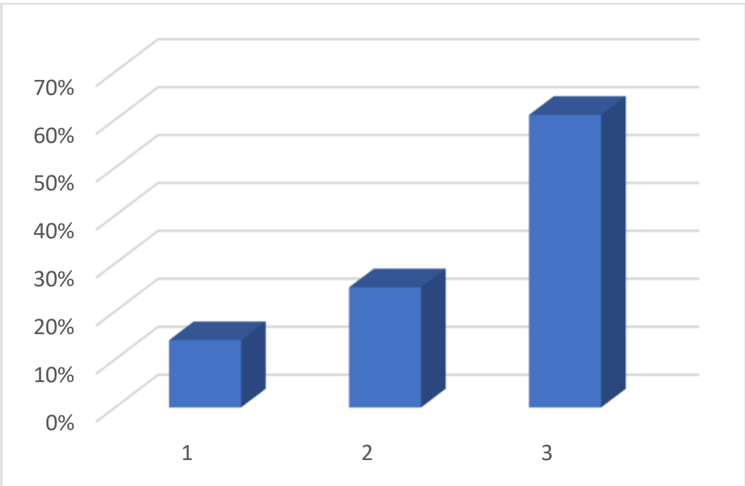

Figure 5. Size of company distribution

Q1. Rank the vision of the digital transformation of your company

1 = (There is no vision)	6%
2 = (Low)	10%
3 = (Medium-high)	31%
4 = (Medium-Low)	29%
5 = (Broad)	23%

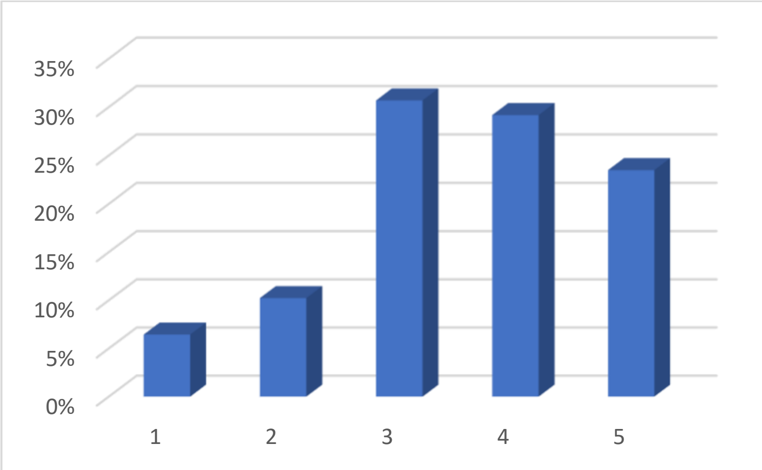

Figure 6. Distribution of company vision of digital transformation rank

Q2. To what extent is Industry 4.0 technologies established and implemented in your company's strategy?

1 = (Not at all)	16%
2 = (Low Level)	17%
3 = (Medium - Low Level)	36%
4 = (Medium - high Level)	26%
5 = (High Level)	6%

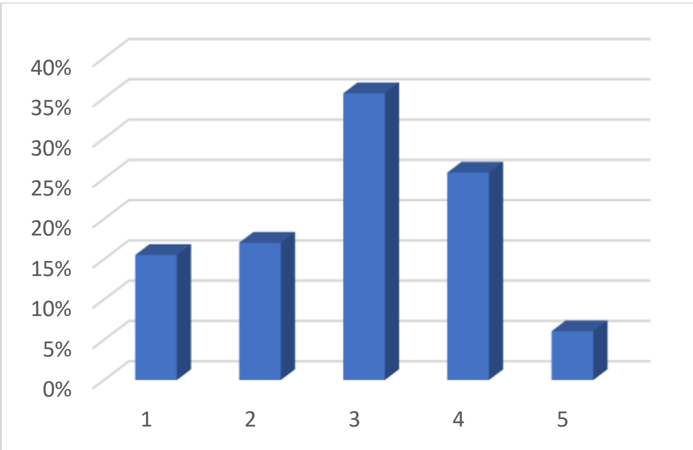

Figure 7. Distribution of Industry 4.0 establishment and implementation

Q3. In the case of no Industry 4.0 technology has been adopted, rank the level of willingness of your company to adopt one or more of it.

1 = (There is no vision)	14%
2 = (Low)	13%
3 = (Medium-Low)	23%
4 = (Medium-high)	40%
5 = (High Level)	9%

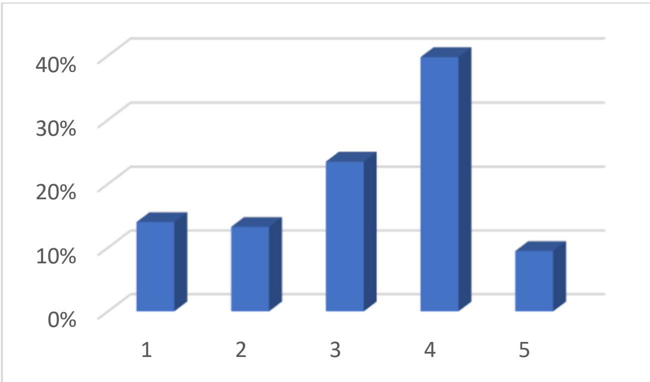

Figure 8. Distribution of willingness to adopt Industry 4.0 technology

Q4. In which parts of your company have you invested in the implementation of industry 4.0 in the past two years?

1. R&D
2. Production/Manufacturing
3. Purchasing
4. Logistics/transportation
5. Marketing & Sales
6. Service
7. IT
8. Human resource management

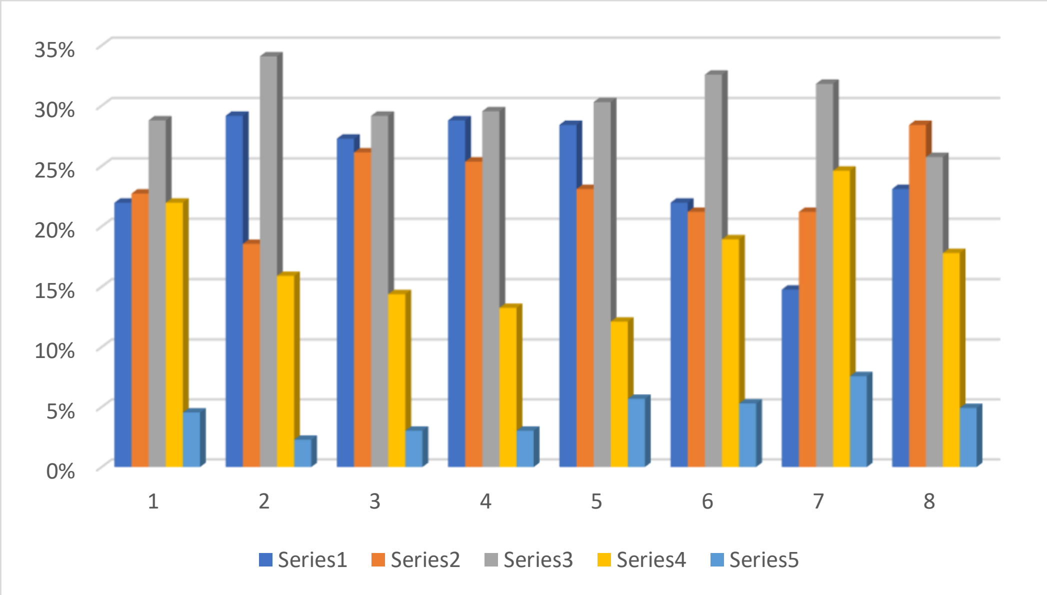

Figure 9. Distribution of company's part that has been invested for Industry 4.0 technologies

Q5. Level of importance of these Industry 4.0 technologies for your organization (Series 1 = No; Series 2 = Basic; Series 3 = Intermediate; 4 Series = High; Series 5 = Advanced)

chnology
1. Virtual/Augmented Reality
2. Robotics/Automation
3. Cyber Physical Systems
4. Collaborative robotics
5. Internet of Things
6. Wireless Technologies
7. Big Data/Data Science
8. Cloud computing
9. Cybersecurity
10. Additive manufacturing (3D printing)
11. Simulation systems
12. Nanotechnology
13. Smart materials

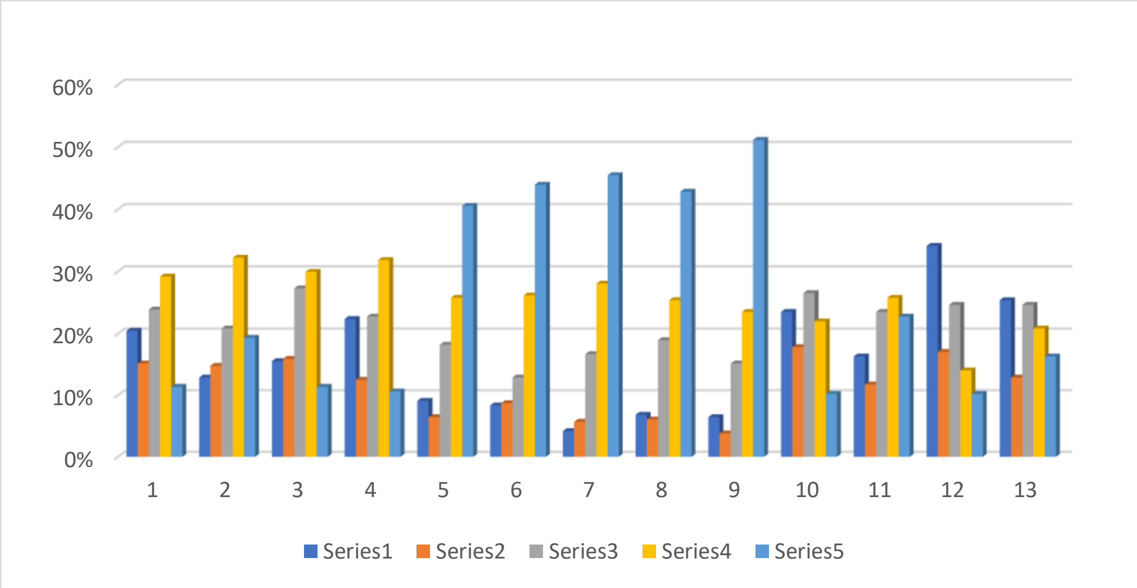

Figure 10. Distribution of importance level of Industry 4.0 in the company

Q6. How would you rate the importance on the following Skills among your employees (Series 1 = not important; Series 2 = less important; Series 3 = just important; Series 4 = very important; Series 5 = very very important)

COMPETENCE
1. Interdisciplinary
2. Team building
3. Leadership
4. Autonomy, responsibility, adaptability, proactivity
5. Fast and focused decision making / problem solving
6. Interpersonal relationship / empathy
7. Intrapersonal relationship / emotional intelligence
8. Ability to work in a group
9. Infographic communication
10. Digital communication

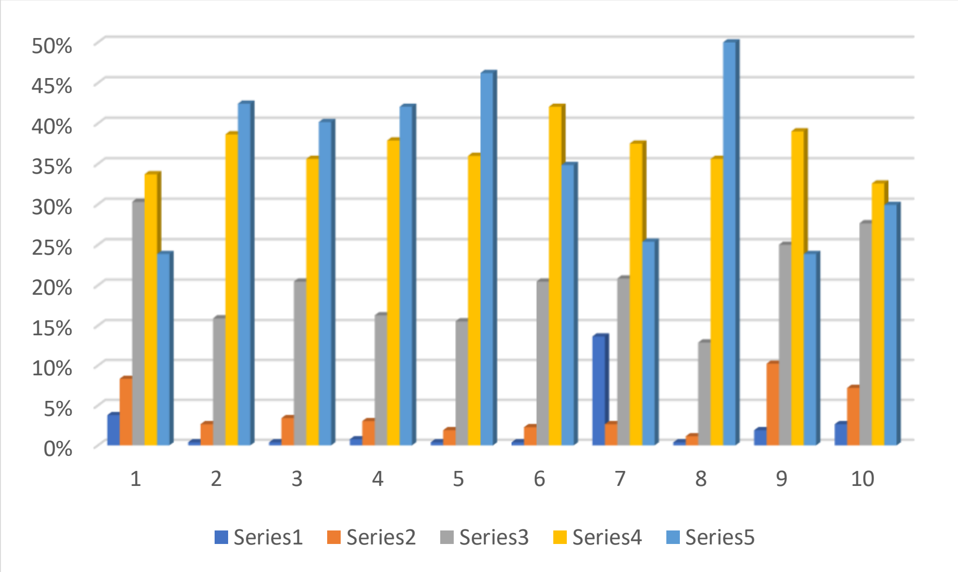

Figure 11. Distribution of importance of skill related to Industry 4.0 among employee

Q7. Express your judgment on the following statements (from 1 = total disagreement to 5 = total agreement)

STATEMENTS
1. Industry 4.0 is not suitable for small businesses
2. Industry 4.0 requires huge investments
3. Industry 4.0 allows large companies to be more agile and therefore threatening for SMEs
4. Industry 4.0 allows SMEs to be more efficient and competitive on the market, thus threatening large companies
5. Industry 4.0 allows product customization that can amplify my competitive strength
6. Industry 4.0 is important but requires skills that I do not have
7. Those who fail to seize the opportunities offered by these innovations risk being excluded from the market

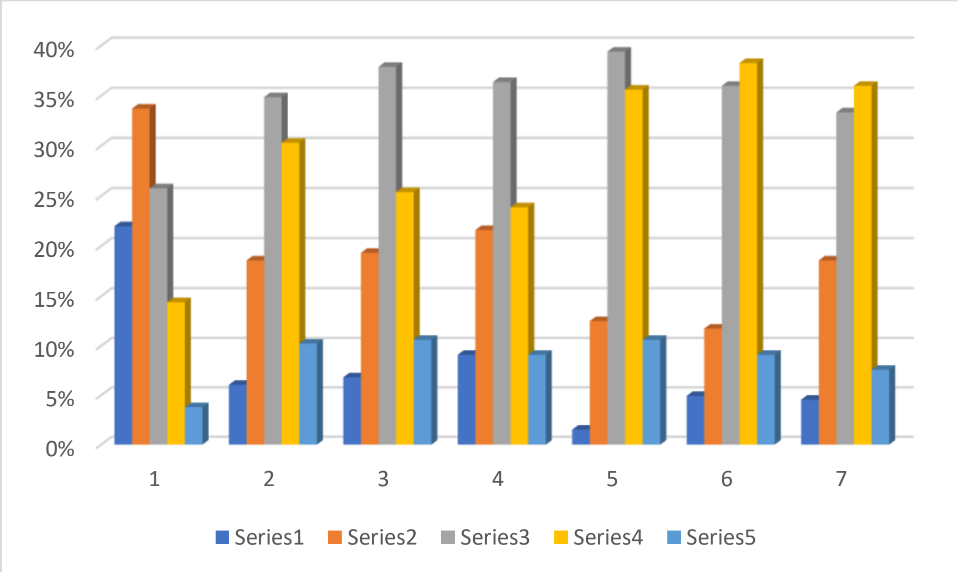

Figure 12. Distribution of judgement towards Industry 4.0

Q8. In which parts of your company have you invested in the implementation of industry 4.0, and what are your plans for the future in the next 5 years?

1. R&D
2. Production/Manufacturing
3. Purchasing
4. Logistics/transportation
5. Marketing & Sales
6. Service
7. IT
8. Human resource management

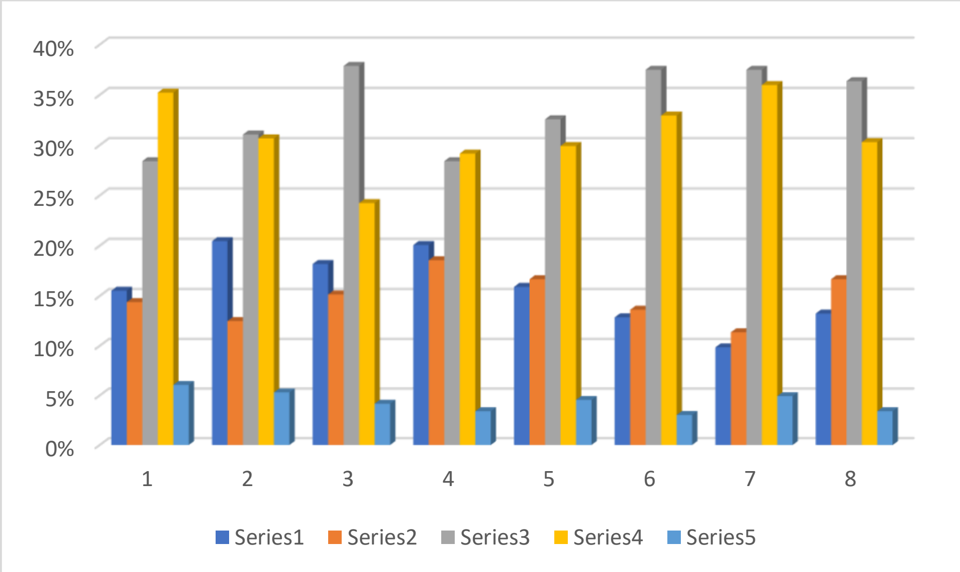

Figure 13. Distribution of parts of company that has been invested and implented with Industry 4.0 technologies

Q9. What is the most frequent technologies you use in wireless networks?

5 = 5G and beyond	16%
4 = IoT	19%
3 = LoRa or NB-IoT	8%
2 = Satellite communications	49%
1 = Not applicable	8%

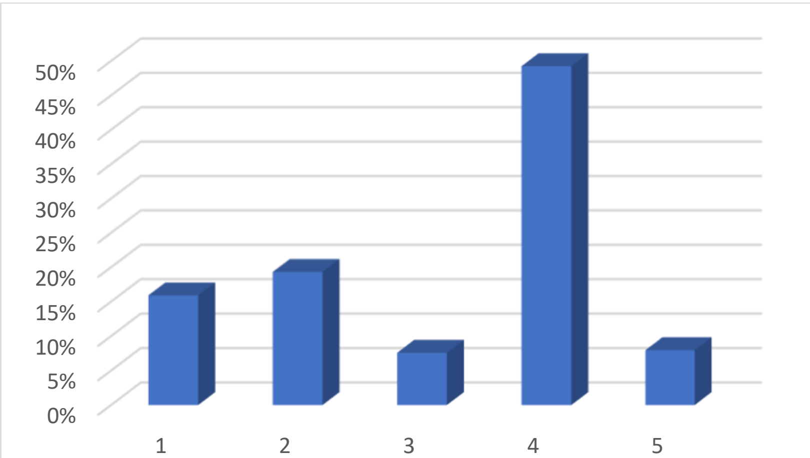

Figure 14. Distribution of most frequently used technologies in wireless network

Q10. Select the type of data your company have.

1. Structured data (e.g., tables)
2. Unstructured data (e.g., image, text)
3. Quasi-structured data (e.g., clicks on website)
4. Semi-structured data (XML)
5. Graph data (e.g., RDF graph)

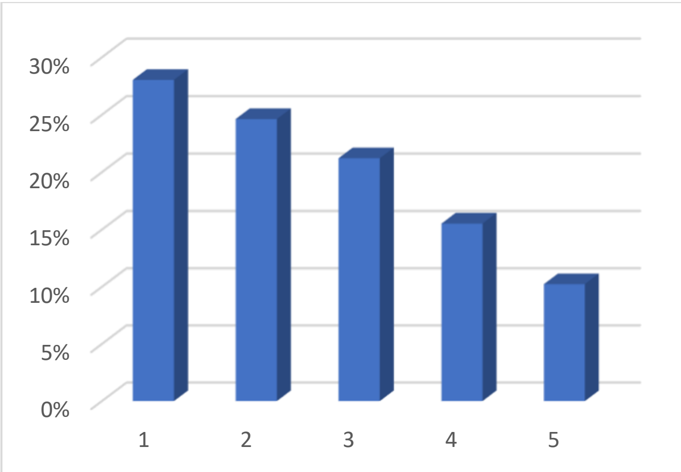

Figure 15. Distribution of data type in company

Q11. What is the programing language your company uses/needs for data engineering and development of analytics?

1. Python
2. Java
3. C++
4. Scala
5. R

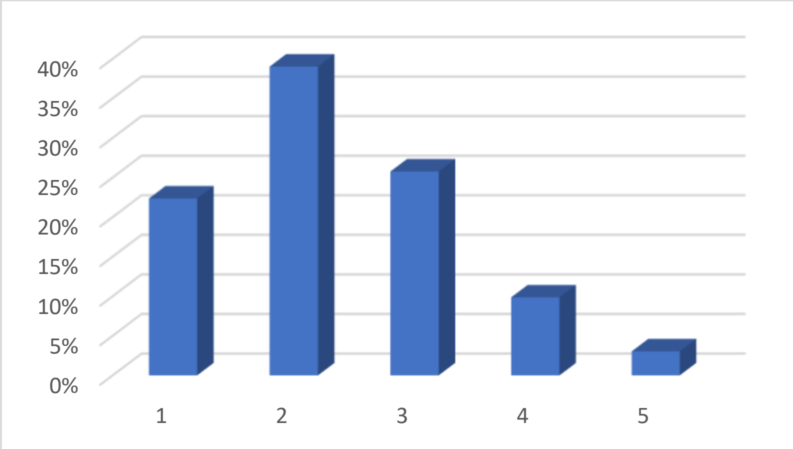

Figure 16. Distribution of programming language used/ needed in company

Q12. What types of data analytics your company uses/needs?

1. Descriptive Analytics (E.g., looking backward, detecting pattern,)
2. Exploratory Analytics (E.g., Spot Anomalies,)
3. Causal/Diagnostic analytics (E.g., Discover a cause or causal relationship)
4. Predictive Analytics (E.g., forward looking, forecast future state relationship,)
5. Prescriptive analytics (E.g., optimal decision for future decision, optimization and decision rules for future events)

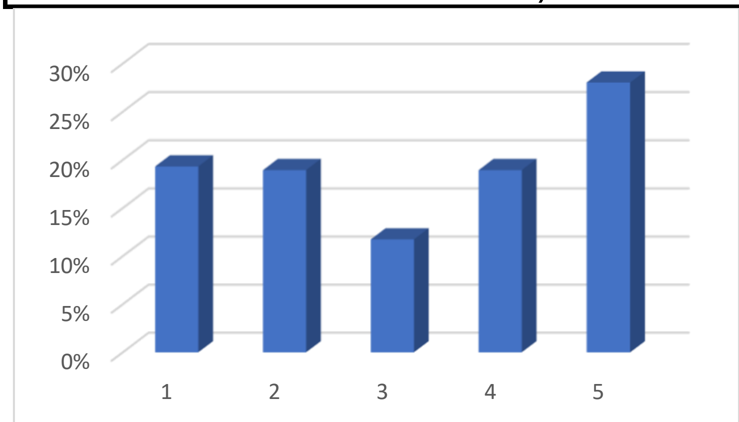

Figure 17. Distribution of usage/ need of data analytics in company

Q13. What are the skills requirement for skills development regarding Industry 4.0 for industrial engineering and management domain? (from 1 = very low to 5 = very important)

ills	
Knowledge about ICT	
a)	Basic information technology knowledge
b)	Ability to use and interact with computes and smart machines like robots, tablets etc.
c)	Understanding machine to machine communication, IT security & data protection
Ability to work with data	
d)	Ability to process and analyze data and information obtained from machines
e)	Understanding visual data output & making decisions
f)	Basic statistical knowledge
Technical know-how	
g)	Inter-disciplinary & generic knowledge about technology
h)	Specialized knowledge about manufacturing activities and processes in place
i)	Technical know-how of machines to carry out maintenance related activities
Personal Skills	
j)	Adaptability and ability to change
k)	Decision making
l)	Working in a team
m)	Communication skills
n)	Mindset change for lifelong learning

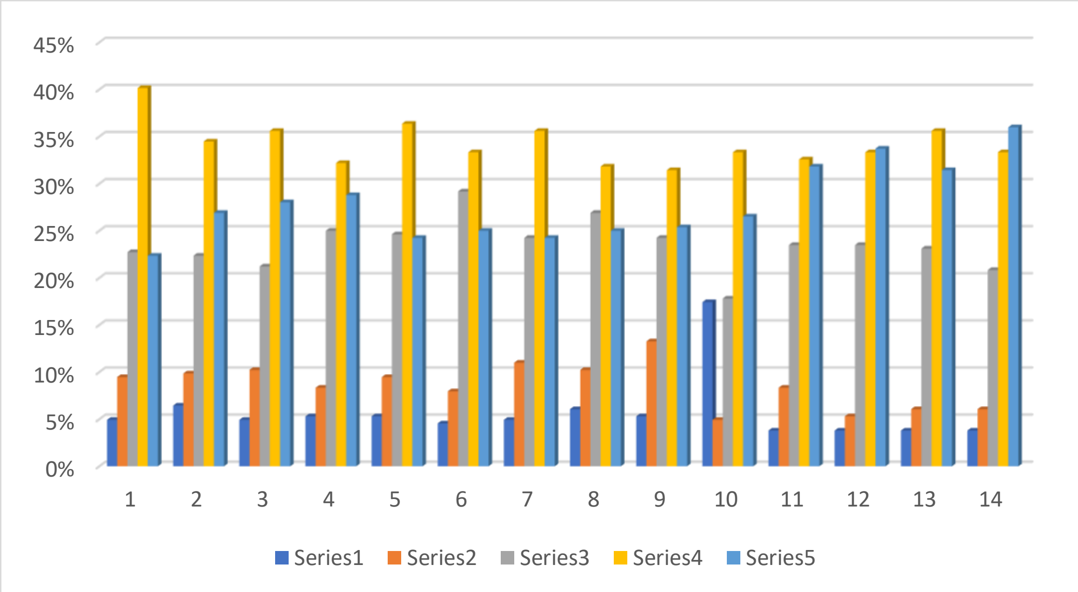

Figure 18. Distribution of skill requirement regarding Industry 4.0 technologies

Q14. What types of Artificial Intelligence (AI) algorithm that your company uses/needs?

a)	Machine Learning
b)	Deep Learning
c)	Convolutional Neural Networks (CNNs)
d)	Recurrent Neural Networks (RNNs)
e)	Reinforcement Learning
f)	Generative Adversarial Network

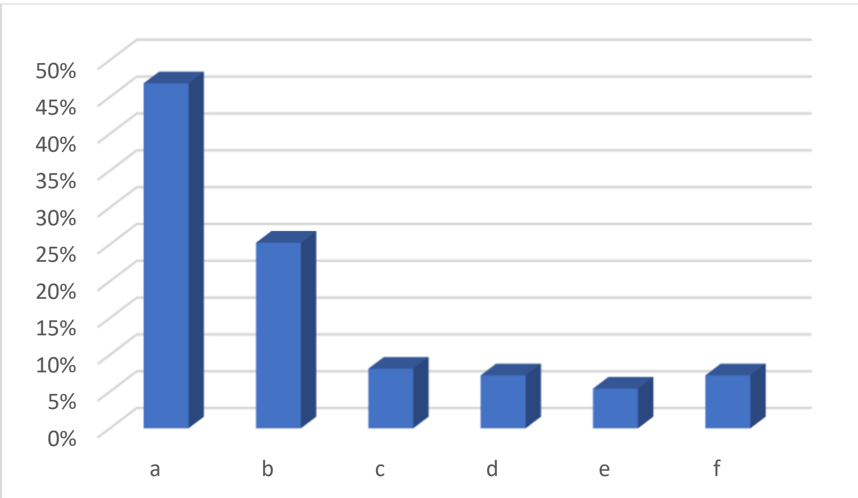

Figure 19. Distribution of AI algorithms usage/ need by company

Q15. Select the type of technology your company uses/needs to implement Artificial Intelligence (AI)?

a) Natural Language Generation	
b) Speech Recognition	
c) Virtual Agents	
d) Machine Learning Platforms	
e) AI-optimized Hardware	
f) Decision Management	
g) Deep Learning Platforms	
h) Biometrics	
i) Robotic Process Automation	
j) Text Analytics and Natural Language Processing (NLP)	
k) Image Recognition	

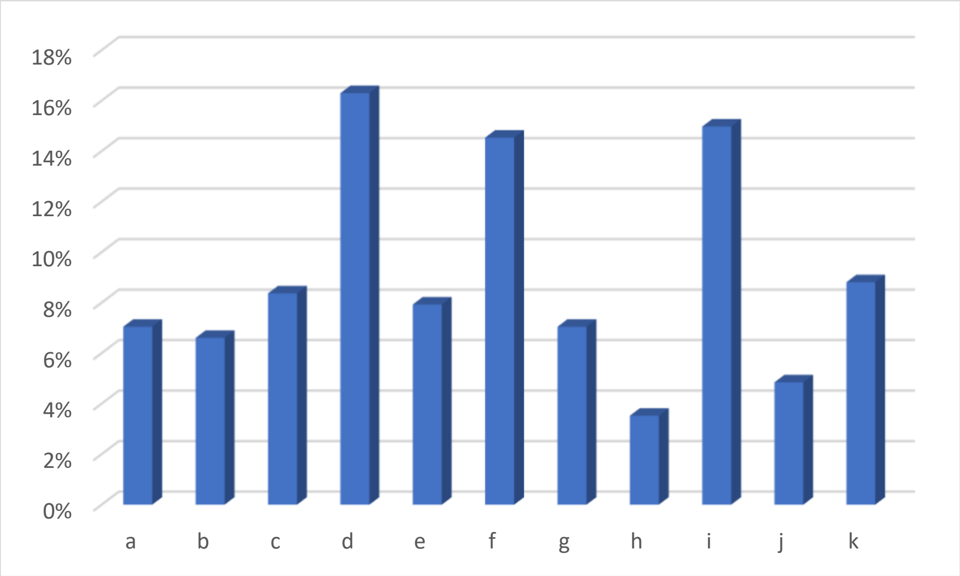

Figure 20. Distribution of technologies type usage/ need to implement AI by company

Q16. What is the programing language your company uses/needs for Artificial Intelligence (AI)?

a) Python
b) Java
c) C++
d) Prolog
e) Lisp

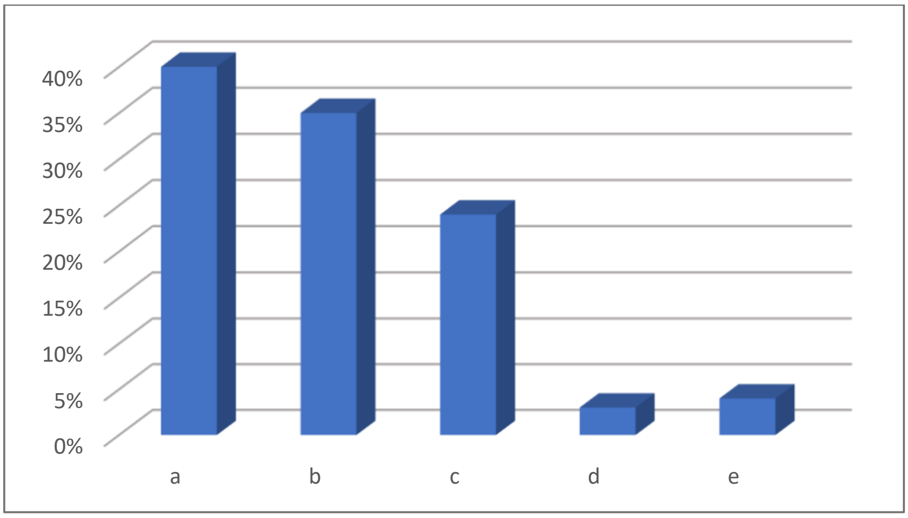

Figure 21. Distribution of programming language usage/ need for AI in company

Q17-Q20. What are the effects of Industry 4.0 and, more generally, of the digitization of manufacturing? (from 1 = total disagreement to 5 = total agreement)

EFFECTS
1. It will increase labor productivity
2. Capital productivity will increase
3. Total factor productivity will increase
4. It will allow you to increase your market share
5. It will allow you to defend your market share
6. It will increase the quality of your products/services
7. It will allow you to place your product in a higher and more profitable area
8. It will allow you to develop different business models

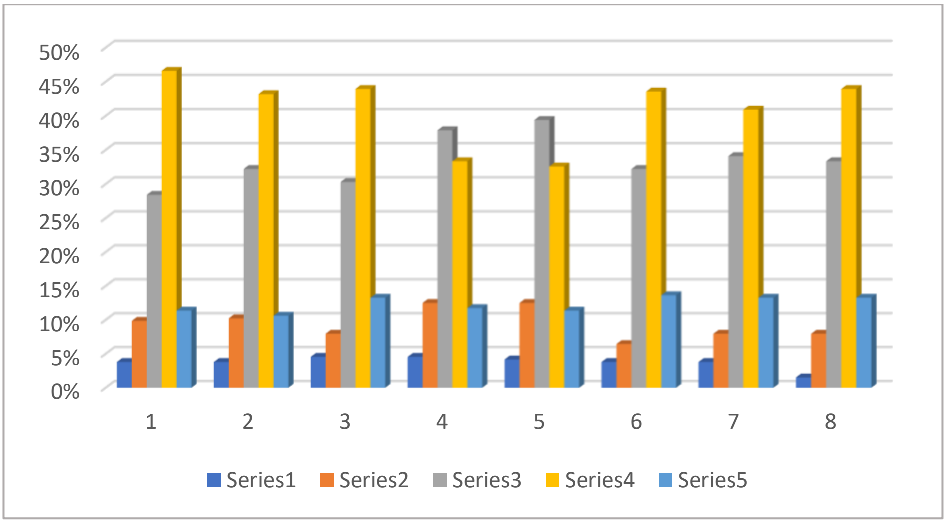

Figure 22. Distribution of effects of Industry 4.0 in company

3.4 Discussion

In the context of the implementation of the level of Industry 4.0, it is possible to assert that 31% of interviewees have a good vision of digital transformation and 29% of them medium-high vision. Furthermore, 36% seems to have a good implementation of Industry 4.0 technologies and, if no Industry 4.0 technology has yet been adopted, there is a medium-high level of willingness of the company to adopt one or more of them. It is interesting to evaluate the investment sectors in the implementation of Industry 4.0 in the last two years, including IT and R&D, while 36% and 30% will invest in the next 5 years in IT and human resource management, respectively.

The most important technologies that have been analyzed for companies were Big Data / Data Science and Cybersecurity. Furthermore, 22% of interviewees totally disagree that Industry 4.0 is not suitable for small companies and only 4% totally agree. On the contrary, we have that 36% agree that Industry 4.0 is important but requires skills that often the company does not have.

The most used technologies in wireless networks were 49% of satellite communications and 28% use prescriptive analytics (E.g., forward looking and forecast future state relationship,) through a language of Java programming at 39%. The most important skills requirements for the skills development were personal skills and, in particular, working in a team with a percentage of 34% and a mindset change for lifelong learning with 36%.

Regarding the implementation of Artificial Intelligence, the technology used at 16% is the Machine Learning platforms, followed by Decision management and Robotic Processes Automation at 15%. The same result is obtained for algorithms where we have 47% of the use of Machine Learning. About the effects of the implementation of Industry 4.0, 47% believe that this will lead to an increase in labor productivity and 44% that there will be an increase in the quality of products and services and in the total factor productivity.

4. Conclusion

Industry 4.0 represents one of the most challenging themes for engineering design and also for engineering education. At this moment there are few studies in the field of engineering teaching that aim to investigate how the educational needs of students and of the industrial workforce are changing. On this basis, this Shyfte 4.0 analysis would like to investigate which are the necessary skills and expertise required to be ready for the implementation of Industry 4.0. In particular, a questionnaire was developed to analyze this situation. It has been administered to Managers and Senior Managers, Chief Executive Officer, Directors, IT Executives and a total of 500 workers who participated in the survey. The questions were aimed to investigate some key issues of Industry 4.0 and digital skills. The collected answers provided a picture of the actual situation in this different country with some relevant considerations about the benefits of Industry 4.0.

Annexes A : References

[1] Arbuckle, J. L., Marcoulides, G. A., & Schumaker, R. E. (1996). Advanced structural modeling: Issues and techniques.

[2] Azevedo, S.G., Ferreira, J. (2007). A model of logistics performance measurement: A structural equation modelling approach. International Journal of Operations and Quantitative Management. Vol. 13(3), pp. 149-171.

[3] Baldwin, B. (1989). A primer in the use and interpretation of structural equation models. Measurement and Evaluation in Counseling and Development, 22, 100–112.

[4] Basl, J. (2017). Pilot study of readiness of czech companies to implement the principles of industry 4.0. Management and Production Engineering Review. Vol.8, pp.3-8.

[5] Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588–606.

[6] Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

[7] Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. Psychological Bulletin, 107, 256–259.

[8] Bollen, K. A., & Long, S. J. (1993). Testing structural equation models. Newbury Park, CA: Sage.

[9] Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 132–162).

[10] Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. Psychological methods, 7(4), 403.

[11] Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research & evaluation, 10(7), 1-9.

[12] Gefen, D., Straub, D., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. Communications of the association for information systems, 4(1), 7.

[13] Gerbert, P., Lorenz, M., Rüßmann, M., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of Productivity and Growth in Manufacturing Industries. BCG Perspectives (bcg.com).

[14] Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). Multivariate data analysis with readings (3rd ed.). New York: Macmillan.

[15] Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 76–99). Thousand Oaks, CA: Sage.

[16] Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351–362.

[17] Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling, 10, 333–351.

[18] Loehlin, J. C. (1987). Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, NJ: Lawrence Erlbaum.

[19] Lu, Y. (2017). Industry 4.0: A Survey on Technologies, Applications and Open Research Issues. Journal of Industrial Information Integration. Vol. (6), pp. 1-10.

[20] MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. Psychological Methods, 11, 19–35.

[21] MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130–149.

[22] Marsh, H. W., Balla, J. R., & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation model- ing: Issues and techniques (pp. 315–353). Mahwah, NJ: Lawrence Erlbaum.

[23] Matteo Mario Savino, Enkhjargal Batbaatar. Investigating the resources for Integrated Management Systems within resource-based and contingency perspective in manufacturing firms. Journal of Cleaner Production 104 (2015) 392-402.

[24] McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model:Noncentrality and goodness

of fit. Psychological Bulletin, 107, 247–255.

[25] Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. Structural Equation Modeling, 7, 36–73.

[26] Mulaik, S. A., James, L. R., Alstine, J. V., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430–445.

[27] Pearl, J. (2009). Causal inference in statistics: An overview. Statistics surveys, 3, 96-146.

[28] Rachna Shah , Susan Meyer Goldstein. Use of structural equation modeling in operations management research: Looking back and forward. Journal of Operations Management 24 (2006) 148–169.

[29] Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 3(4), 369-379.

[30] Sadia R., Bekhor S., Polus A. (2017) Structural equations modelling of drivers' speed selection using environmental, driver, and risk factors, Accident Analysis & Prevention, In Press, Corrected Proof

[31] Savino, M.M., Shafiq, M.. (2018). An extensive study to assess the sustainability drivers of production performances using a resource-based view and contingency analysis, Journal of Cleaner Production Vol. 204, pp. 744-752

[32] Schmidt, R., Möhring, M., Härting, R.C., Reichstein, C., Neumaier, P., Jozinović, P. (2015). Industry 4.0 - Potentials for Creating Smart Products: Empirical Research Results 18th International Conference on Business Information Systems, Lecture Notes in Business Information Processing (LNBIP). Vol. 208

[33] Shumow, L., & Lomax, R. (2002). Parental efficacy: Predictor of parenting behavior and adolescent outcomes. Parenting: Science and practice, 2(2), 127-150.

[34] Shumow, L., & Lomax, R. (2002). Parental efficacy: Predictor of parenting behavior and adolescent outcomes. Parenting: Science and practice, 2(2), 127-150.

[35] Srivastava, S., Dubey, R. (2014). Supply chain skill gap modelling using interpretive structural modelling and MICMAC analysis. International Journal of Operations and Quantitative Management. Vol. 20(1), pp. 33-47.

[36] Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. Structural Equation Modeling, 7(2), 149-162.

http://www.shyfte.eu/