The dominant factors driving social sustainability performance of global supply chain: An exploratory study using machine learning approach *(work in progress)*

Despite the growing awareness of sustainability especially in the global supply chain in the business context, there are few empirical studies on the social dimension of sustainability. Moreover, the complexity to manage social sustainability performance in the supply chain has put business reputation at risk beyond their ordinary business operation. To manage the social sustainability risk in the supply chain, business need to understand more on what factors drive social sustainability performance of their supply chain. By identifying the driving factors of the social sustainability performance, businesses will be able to mitigate the inherent risk and manage their supply chain. Understanding the rarity of the subject on social dimension, complexity of supply chain, and urgency to mitigate the risk, the purpose of the study is to explore the possible factors that dominantly driving the producer's social sustainability performance in the context of global supply chain with their buyer's.

An embedded case study has been chosen to understand the phenomena within the context of business engagement of buyer's & their supply chain producer's. The scope of the topic is on exploring possible driving factors either on the producer's side with 27 factors, next is on the buyer's side with 3 factors and last is on the macro environment with 7 factors that influenced the social sustainability performance from 2015 to 2016. An international sustainability supply chain organization based in europe having 2.000 buyer's & 50.000 producer's has been selected due to global coverage, reliability and massive dataset and relevancy to the research subject. The study leverage a large sample of independent audit dataset collected by the social auditor to each producers within certain time-frame to be reported to the producer's linked buyer's. The social sustainability performance measure in the study is the aggregate measure of thirteen social sustainability performance area such as the rights of freedom of association and collective bargaining; no discrimination; fair remuneration; decent working hours; occupational health and safety; no child labour; special protection for young workers; no precarious employment; no bonded labour; protection of the environment; and ethical business behaviour.

To cope with complexity of the context, the study use an exploratory data mining analysis techniques from the machine learning domain to explore dominant factors from multiple possible factors from different view & at the same timeframe altogether. The dataset are analysed using decision tree ensemble based on random forest algorithm to reduce the randomness & variability of the result. A supervised machine learning approach conducted to find an optimal algorithm to predict the future outcome based on accuracy. The factors then rank based on their importance to the social performance to find the dominant factors. Furthermore, the cluster analysis is use to find un-observe or hidden forces base on the factors. The need to mitigate the reputational risk drives the case study to have three unit of analysis of the producer's that are linked based on the buyer's business model based on relationship type, business-to-business and business-to-consumer. Comparative analysis of these different type of buyer's business model might affect the behavior of the producer's & might affect the dominant factors.

The study finds seven driving factors are dominantly influenced producer's social sustainability supply chain performance with prediction accuracy of 75%. The rest of 30 factors only increase additional prediction accuracy of 7% to 82%, suggesting the less importance of these other factors. These dominant factors are first, the rate between workers legal minimum wage and average living cost around the producer's factory. Second is the

proportion of women workers as compared to total workers in the producer's factory. Third is the absolute number of workers in the producer's factory. Fourth is the average annual financial turnover of buyers that are doing business with the producer'. Fifth is the annual financial turnover of producers. Sixth is the proportion of number of management as compared to number of worker in the producer's factory. The seventh & the last is the rate between workers remuneration & average living cost around the producer's factory.

Within these dominant factors there are four identifiable forces such as labor-economics, gender, management & trade. Labour-economic force is where economic of the labor such as legal minimum wage, remuneration could cope with their corresponding living cost and being represented by the factors one & seven. Gender force is where women proportion in the workforce and being represent by factors two. Business-structure force is where number of worker required, their management proportion to their worker & economics capabilities to produce the product and being represented by the factors three, five & six. Buyer's economic forces is where buyer's economic turnover linked to the producer's interest to work in the supply chain relationship & being represented by factors four. It is then condensed into the forces of labour, economic, management, gender & trade.

Furthermore, these dominant factors & their forces are consistently present for each different type of buyer's business model that are linked to their producer's. Whether the producer's is specifically linked to either to the buyer's type B2B and B2C or producer's who serve both B2B & B2C. The differences are only in the ranking for each dominant factors on each business model type. It suggest that regardless the supply chain relationship of the producer's and their buyer's based on the type, the seven dominant factors still prevalent and consistent.

The exploratory nature of the study due to its rarity has served its objective for further study by more understanding on social sustainability in the supply chain. Although the result of predicting the social performance based on the seven dominant factors produce high accuracy, there might be other dominant factors outside the study that might be important to the social sustainability performance. The difficulties to manage sustainable supply chain in particular on social dimension has been demystify by the study. Businesses, government and organization interested in sustainability should be able to understand more about the core focus of factors to tackle social sustainability performance of the supply chain. The study has open new opportunities to improve social life of million workers around the world through producers who produce goods and their corresponding consumers through buyers who deliver goods. The study open new path to understand complexity of the social sustainability of global supply chain by considering large number of factors in the same context at the same time with new method in machine learning. It open the possibility of new & development of theory to understand more on the sustainability of global the supply chain.