TITLE

Garini F., Hansstein F., Nigri A., Stasi A.

May 2, 2021

Abstract

To do...

Keywords:

1 Introduction

In recent years, the market of ready-to-eat (hereafter RTE) salads has become a healthy trend among consumers worldwide. While RTE salads are popular among Western consumers, the market is still expanding in the Middle Kingdom. In fact, lifestyle in urban China has dramatically changed in the last years. Most consumers, especially young and urban, pay attention to their health, diet, and physical appearance much more than in the past. At the same time, they conduct a busy and fast-paced life, always trying to save time and money. Health and future uncertainty have become more prominent after the Covid-19 crisis, as it emerged from recent data (Daxue Consulting, 2021). Eating more fruits and vegetables is associated with a robust immune system and risk reduction of many cardiovascular diseases and cancer (A. Gibson et al., 2012; Gu et al., 2021). In fact, attributes of RTE salads such as: healthy, easy-to-find, and time and money efficient, let us reasonably assume that this consumption pattern will keep growing in China. However, the intake of vegetables in China is still below the level recommended by the Chinese dietary guidelines (Xiao et al., 2015). Therefore, encouraging the consumption of vegetables, including fresh-cut products, can be a successful strategy to improve the population's health status. Literature presents plenty of evidence about food consumers' preferences. It shows that they are oriented towards quality, safety, nutritional elements, freshness, richness of assortment, etc (Santeramo et al., 2017). In addition, other widespread findings state that medium-high income consumer target, generally also high educated people (middle-class), is interested in purchasing products with a higher level of service and quality (Artés et al., 2009; Watada et al., 1999; Soliva-Fortuny et al., 2002; Rico et al., 2007). Meanwhile, contributions on RTE fruits and vegetables (hereafter FV) are still limited. However, common outcomes regard income as a driver of purchase (Marshall et al, 1994; Cassady et al., 2007) as well as time pressure, social environment and eating-out habits (Frewer et al. 2001; Buckley et al., 2007). Other papers highlight that quality, safety and labeled information are also important (Pilone et al., 2017; Baselice et al., 2017; Santeramo et al., 2017; Pollard et al., 2002; Baselice et al., 2014). All these contributions, however, present empirical evidence of Western countries. To the knowledge of the authors, no one ever proposed an empirical study based on Chinese consumers' data, despite the preeminent growing expectations on the specific market. Given these premises, the first objective of the paper is to offer a preliminary contribution to the understanding of the factors affecting the consumption of ready-to-eat salads in China. Consumers' behavior studies have most of the time referred to econometrics in order to predict choices (cit. . . .). However, limited data associated with a high number of variables/factors have always been an issue in order to have easy interpretation and good prediction ability of models for the inference process (cit. . . .). In recent years, however, different modelling approaches have been brought to light, also because of the remarkable increase in computational power (cit. . . .). Machine learning (hereafter ML) approaches, in fact, could present advantages over traditional econometric models in certain situations. It is enormously useful in our context where a large number of factors may affect the fresh-cut customer's choice. Traditional techniques, such as the logistic regression, may exhibit a rigid structure frequently leading to a functional misspecification. Instead, the "distribution-free" approach does not impose a probabilistic function, therefore, the relationship among variables is directly generated from data (cit. . . .).

To the best of our knowledge, as literature lacks fresh-cut consumers' behavior study, also it is likely to have a limited number of studies that use ML models for food preferences predictions. As a consequence, the second objective of the paper is to propose, while presenting the results of the empirical analysis, a comparison among traditional Logit, random forests (RF), Deep Neural Network (DNN), Regression Tree (RT) in a food preferences analysis. Finally, the paper proposes manifold contributions: (i) studying a never explored market in order to provide foresights to investors and policy makers; (ii) testing competing models in order to reduce prediction errors; (iii) bringings new perspectives and findings to the literature of the fresh-cut; (iv) contributing to the nascent literature on using ML methods to predict customers' behaviors.

2 Framework

FAO statistics about food balance highlight that, while western countries show decreasing consumption of Fruits and Vegetables (hereafter F&V), the rest of the world, such as Asian countries, show the opposite. In fact, population growth, increasing trade, nutrition awareness, and migration are all factors having an incidence on those trends (FAO, 2017). Despite the general level of consumption regarding FV, when looking at population status, some other evidence of societal emergencies comes to light. Malnutrition comes in a number of forms that not only affects a person's health and well-being, but also places heavy burdens on families, communities and states (FAO and WHO, 2014). Ending hunger, achieving food security and improving nutrition are all key steps toward sustainable development (UN, 2016). Food safety is also a key concern, as unsafe food remains a major cause of disease and death (WHO, 2015). Meanwhile, changes in dietary patterns around the world have consequences for public health and sustainable development. 'Triple burden' of malnutrition remains a global health emergency. The 'triple burden' of malnutrition weighing on most countries consists of under nutrition, micronutrient deficiencies, and overweight and obesity (FAO, 2017). In the last decades, the consumption of more nutritious foods increased worldwide, however, relevant trends differ across regions. For instance, fruit consumption raised in all areas, but vegetable consumption increase concerns a more limited number of countries (FAO, 2017). When correlating consumption and income, typically, per capita fruit intake tends to rise with income, whereas vegetable consumption declines (FAO, 2017). In fact, given the great concern about contemporary dietary habits, governments in several countries have launched informational and educational initiatives aimed at increasing public awareness about the benefits of FV. As a consequence, the effectiveness of this campaign is evaluated by several authors (Seiders and Petty, 2004; Gordon et al., 2006; Mazzocchi et al., 2009). Food policy and regulatory guidelines are prime tools to achieve a healthy diet in a population, with a remarkable impact on the health care systems and social and public health policies. In this regard, China has a long-standing commitment to policy guidance on food security-related nutrition. Since the '90s it started with the "China Nutrition Improvement Action Plan (1996-2000)", the "Management Measures for Nutrition Improvement (2010)" and the "National Nutrition Plan (2017–2030)" in 2017, which aims to improve nutrition and health on a national level and set national nutrition strategies for 2020 and 2030. Recently, China made great strides in food security and nutrition, implementing the "Healthy China Action Plan (2019–2030)" (HCAP). This regulation centralizes efforts to improve the populations' nutrition status, facing many challenges, especially obesity and chronic diseases.

Indeed, aiming to increase the daily consumption of FV, as recommended by the World Health Organization (World Health Organization (2008)), the "fresh-cut" sector is playing an important role due to the convenience (in terms of time saving for washing and preparation) and freshness of this minimally processed product. In line with the definition given by the International Fresh-Cut Produce Association (IFPA), fresh-cut FV are minimally processed products; that are washed, cut, mixed, and packed. Since their first appearance in Europe in the early 1980s, they have become more and more popular in consumers' market baskets because of time convenience, quality and safety attributes that are generally valued positively by consumers (Artes et al., 2019). Those products are also considered as ready-to-eat products (hereafter RTE), that is "food intended by the producer or the manufacturer for direct human consumption without the need for cooking or other processing to eliminate or reduce to acceptable level micro-organisms of concerns" (Regulation 2073/2005/EC). In China, RTE products regulation is included in the Food Safety Law of 2015 (HFG - Law and Intellectual Property, 2016). Given the risk of raw food contamination with dangerous bacteria, RTE food safety is currently the main concern for authorities. Therefore, encouraging proper handling and storage to avoid foodborne diseases is a priority for public and private food institutions. RTE products include fresh-cut fruits and vegetables and other ready-to-eat meals like cereals, dairy, cold cuts, and meat. The nutritional benefits of RTE food have been extensively acknowledged by experts (Santeramo et al., 2018; Singla et al., 2020), despite some warns about health risks associated with high sugar and fat content (Poti et al., 2016; Thike et al., 2020). On top of the market potentiality of RTE meals, there is the provision of a balanced diet in terms of nutrient intake, without giving up on taste and quality, both for kids and adults (A. A. Gibson Partridge, 2019). The success of RTE fresh meals in Western countries is a sign of how innovation and integration between traditional agriculture and the food industry can meet the needs of modern consumers (Brennan et al., 2013; Massaglia et al., 2019) RTE salads are thus a food product suitable for combining the needs of modern consumers.

...

...

Descrizione ML

ML has advantages over traditional econometric models in certain situations. It is enormously useful in our context where a large number of factors may affect the fresh-cut customer's choice. Canonical inference tools, as in the case of logistic regression, may exhibit a rigid structure, leading to functional misspecification. On the contrary, the so-called "distribution-free" approaches do not impose a probabilistic assumption, therefore the functional relationship between variables is directly approximated from the data. Although some ML algorithms use hyperparameters, such as the number of trees and learning rates, these hyperparameters come by a fine-tuning procedure, thus are optimally chosen over a broad grid, Despite is widely recognized that ML suffers from an interpretation issue, it fits our purpose of predicting the customer's behavior, particularly because the data are large and predictors are convoluted. Given the emerge of ML in prediction tasks for business, we expect the method to generate practical directions for our study. To the best of our knowledge, researchers have not explored

ML models for food fresh-cut consumers' predictions. As result, we apply random forests (RF) for predicting access to Chinese fresh-cut consumers' predictions, integrating the analysis with Deep Neural Network (DNN), Regression Tree (RT)... We selected these ML models over others because they scale with the volume of information without damaging statistical efficiency, versatile than others, and typically demonstrate good predictive accuracy as they are robust to outliers.

Conclusioni

The approach we follow is peculiar and new, providing manifold contributions. Indeed, (i) we bring new perspectives and findings of the fresh-cut products among Chinese customers, where the literature lacks a focus on this topic. (ii) Aiming at reducing the prediction errors, we contribute to the nascent literature on using ML methods to predict customers' behaviors. Thereby, this work aims at providing new insights to the analysis and prediction of fresh-cut customers' behavior in China even using recent innovations, offering peculiar information for both policymakers and practitioners.

3 Models

3.1 Logit

To explore the role of specific factors affecting the consumption of ready-to-eat salads in China, we firstly run a logit model. Logit models are primarily used in discrete choice analysis and are helpful to assess the odds ratio of an outcome exposure or food choice. In consumer behavior and marketing research, logit models have been widely used to understand purchasing behavior (see for example Nevo, 2011; Roberts Lilien, 1993).

Our independent variable can be treated as a realization of a random variable that follows a binomial distribution $Y_i \sim B\left(n_i, \pi_i\right)$. Thus, we define generalized linear model with binomial response and link logit as:

$$\operatorname{logit}\left(\pi_i\right)=\mathbf{x}_i^{\prime} \boldsymbol{\beta} \tag{1}$$

where the logit of the underlying probability π_i is a linear function of the predictors where $\mathbf{x}_i$ is a vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. Thereby, β_j represents the change in the logit of the probability associated with a unit change in the j -th predictor holding all other predictors constant. Exponentiating Equation 1 we find that the odds for the i -th unit are given by

$$\frac{\pi_i}{1-\pi_i}=\exp \left\{\mathbf{x}_i^{\prime} \boldsymbol{\beta}\right\} \tag{2}$$

The related log-likelihood function is defined as:

$$\log L(\boldsymbol{\beta})=\sum\left\{y_i \log \left(\pi_i\right)+\left(n_i-y_i\right) \log \left(1-\pi_i\right)\right\}, \tag{3}$$

where π_i depends on the covariates $\mathbf{x}_i$ and a vector of p parameters $\boldsymbol{\beta}$ through the logit transformation of Equation 1 .

In our specification, the dependent variable indicated if the respondent regularly eats fresh-cut salads. The explanatory variables were age, income, seniority, household size, knowledge, fitness, source of information regarding RTE products, purchase location, and whether consumption occurs for snacking or for lunch. The logit model was conducted on 70% of the original sample as required by predictive statistics.

3.2 Machine Learning

In this section, we deal with the wide framework of unsupervised learning which embraces various models such as Support Vector Machine, Neural Networks and Tree-based Models, used for both regression and classification problems.

According the statistical learning theory (SLT) the problem of supervised learning is formulated as follows. Given a set of 1 training data $D = \{(\mathbf{x}_1, \mathrm{y}_1) \ldots (\mathbf{x}_1, \mathrm{y}_1)\}$ in $\mathrm{R}^{\mathrm{n}} \times \mathrm{R}$ sampled according to unknown probability distribution $\mathrm{P}(\mathbf{x}, \mathrm{y})$, and a loss function $\mathrm{L}(\mathrm{y}, \mathrm{f}(\mathbf{x}))$ that measures the error done when, for a given $\mathbf{x}, \mathrm{f}(\mathbf{x})$ is "predicted" instead of the actual value y.

It is worth pointing out that there is no information on the underlying joint probability functions. Therefore, it is necessary to perform a "distribution-free" approach, where the only information available is a training dataset.

The problem consists in finding a function f that minimizes the expectation of the error on new data, that is, find a function f that minimizes the expected error:

$$\int \mathrm{L}(\mathrm{y}, \mathrm{f}(\mathbf{x})) \mathrm{P}(\mathbf{x}, \mathrm{y}) d\mathbf{x} d\mathrm{y}$$

Since $\mathrm{P}(\mathbf{x}, \mathrm{y})$ in unknown, we need to use some induction principle in order to infer from the 1 available training examples a function that minimizes the expected error. The principle used is Empirical Risk Minimization (ERM) over a set of possible functions, called hypothesis space. Formally this can be written as minimizing the empirical error:

$$\frac{1}{l} \sum_{\mathrm{i}=1}^{l} \mathrm{L}\left(\mathrm{y}_{\mathrm{i}}, \mathrm{f}\left(\mathbf{x}_{\mathrm{i}}\right)\right)$$

Machine Learning tools are the so-called “nonparametric” models. “Nonparametric” does not mean that the ML models do not have parameters at all. On the contrary, their “learning” is the crucial issue here, indeed, unlike in classic statistical inference, the parameters are not predefined and their number depends on the training data used.

3.3 DNN

The term NN refers to a mathematical model inspired by the human brain. It allows computational models composed of multiple processing layers to learn representations of data characterized by multiple levels of abstraction, solving even the most complex problems. Its architecture includes neurons, synaptic connections that link the neurons, and learning algorithms (backpropagation). NN works as a weighted regression in which each unit gets “weighted” information through synaptic links from the other connected ones, returning an output by using an activation function that transforms the weighted sum of input signals. Albeit DL embodies a broad variety of networks differing from each other by the architectural structure, in this section the basic scheme of NN has been presented, the so-called feedforward structure. NN training involves an unconstrained optimization problem where the aim is to minimize a function in high dimensional space the so-called loss function, that measures the difference between the predicted values and observed ones. The back-propagation is the most used algorithm for the training of NNs. The algorithm compares the predicted values against the desired ones (objective) and modifies the synaptic weights by back-propagating the gradient of the loss function. Schematically, the procedure alternates forward and backward propagation steps:

- in the forward step, the prediction is computed fixing the synaptic weights,
- in the backward step, the weights are adjusted in order to reduce the error of the network.

The NN iteratively performs forward and backward propagation and modifies the weights to find the combination that minimizes the loss function $\mathcal{L}$.

3.4 SVM

Support Vector Machines (Vapnik) originated as an altertinative to Neural Networks in the wide context of statistical learning, offering a solution to this trade-off by arbitrarily fixing model accuracy and minimising model complexity; in this way it is able to manage overfiting.
In a classification problem, this translates in an optimal solution identified by a linear hyperplane which correctly classifies observed data and lies as far as possible from them.

3.5 Random Forest

Since Regression Trees usually produces low bias and high variance estimations, they became a good candidate for ensemble methods. Indeed, Random forests basically consist in building an ensemble of decision trees grown from a randomized variant of the tree, this method is useful to get the error reduction pulling down the prediction variance, preserving the bias Starting from a single learning set, the basic idea is to introduce a random perturbation into the learning procedure in order to introduce a differentiation among the trees and combine the predictions of all these trees using aggregation techniques. Breiman **?** proposed a first aggregation method the so-called bagging in which the different trees, are built by using random bootstrap copies of the original data. Its natural evolution, the random forests, has been developed by the same author in 2001 **?**. In the random forests the bagging approach has been extended and combined with randomization of the input variables that are used when considering candidate variables to split internal nodes t. In particular, instead of looking for the best split s^* among all variables, the algorithm chooses a random subset of K variables for each node and then determines the best split using these variables.

3.6 Performance assessment Cross-validation

In a k -fold cross-validation (CV), the original dataset is randomly partitioned into k subsets of approximately equal sizes. At each of the k CV iterations, one of the folds is chosen as the test set, while the $k-1$ others are used for training. The considered performance measure is computed based on the test set. After the k iterations, the performances are finally averaged over the iterations. In our study, we perform 10 repetitions of stratified 5 -fold CV, as commonly recommended (Bischl B, Mersmann O, Trautmann H, Weihs C. Resampling methods for meta-model validation with recommendations for evolutionary computation. Evol Comput. 2012;20(2):249–75) . In the stratified version of the CV, the folds are chosen such that the class frequencies are approximately the same in all folds. The stratified version is chosen mainly to avoid problems with strongly imbalanced datasets occurring when all observations of a rare class are included in the same fold. By "10 repetitions", we mean that the whole CV procedure is repeated for 10 random partitions into k folds with the aim to provide more stable estimates.

3.7 Accuracy Prediction

For binary target variables, we evaluate the level of accuracy and its 95% confidence interval (CI), true positive rate (Sensitivity), true negative rate (Specificity), and Cohen's Kappa (Cohen, 1960). We define $y=1$ for a regular consuption of RTE and 0 otherwise. Then a 2×2 confusion matrix has elements $a_{\text{row,column}}$ with predicted conditions $\widehat{y}=\{1,0\}$ on rows and true conditions $y=\{1,0\}$ on columns. The statistics are defined by,

$$\text{Accuracy } = \frac{a_{11}+a_{22}}{a_{11}+a_{12}+a_{21}+a_{22}} \tag{4}$$

$$
\begin{aligned}
&\text{Sensitivity or True Positive Rate (TPR)} = \tfrac{a_{11}}{a_{11}+a_{12}} \\
&\text{Specificity or True Negative Rate (TNR)} = \tfrac{a_{22}}{a_{21}+a_{22}}
\end{aligned} \tag{5}
$$

We calculate accuracy that refers to the portion of customers correctly classified with respect to RTE regular consumption. We calculate the 95% confidence interval using accuracy's standard deviation generated through iterations. Sensitivity refers to the proportion of regular RTE consumers correctly identified as such. Poor sensitivity implies a large number of inclusion errors, i.e. identifying RTE consumers when in fact they are not. Specificity refers to the proportion of customers correctly predicted to be occasional RTE consumers. Poor specificity implies a large number of exclusion errors. Cohen's Kappa statistic measures the agreement for categorical variable relative to what would be expected by chance. That is,

$$
\begin{aligned}
&\text{Kappa} = 1 - \tfrac{1-\text{ Accuracy}}{1-\text{EP}} \\
&\text{EP} = \tfrac{(a_{11}+a_{12})(a_{11}+a_{21})+(a_{21}+a_{22})(a_{12}+a_{22})}{(a_{11}+a_{12}+a_{21}+a_{22})^2}
\end{aligned} \tag{6}
$$

where, accuracy of prediction is the observed probability, and EP is the probability of random agreement or expected probability. Cohen's Kappa becomes zero if there is no agreement among the predicted and observed response other than what would be expected by chance.

4 Epirical Analisys

4.1 Data

This cross-sectional study examined the consumption of ready-to-eat salads among a sample of Chinese consumers. The consumption of prepackaged fruits and vegetables has increased in China in the last years, especially among young urban workers. Fresh salads are a versatile, healthy, and convenient meal that can easily substitute or accompany a lunch, a snack, or a dinner. A total of 509 questionnaires were collected by trained interviewers between January and March 2019. The study employed a convenient sample, and participants took part in the survey voluntarily. The questionnaire was part of a larger project that collected information on ready-to-eat fruits and vegetables. It was designed following the literature on the consumption of fruits and vegetables and included several sections. Questions included in the analysis are available in the Appendix. The project's coordinators built the first version in English, and the questionnaire was later translated into Chinese by a bilingual speaker. Researchers conducted a questionnaire validation through a pilot test on a smaller sample of potential respondents to make sure that Chinese speakers correctly understood all the questions. The core analysis included behavioral questions on consumption of ready-to-eat-salads, socio-demographic variables, knowledge of fresh-cut products, social norms, and other questions on respondents' habits. All variables included in the model were categorical.

Variable transformation Most variables did not require any modification, with the exception of two indicators that were built after data collection. The first indicator is "knowledge," a variable measuring respondent's level of awareness towards packaged ready-to-eat fruits and vegetables compared with their regular alternative. Respondents had to indicate whether six statements were true or false. They were: "RTE products must be washed before being consumed," "RTE products are more treated," "RTE products are healthier," "RTE products are industrial products," and "RTE products are more controlled and selected." If the respondent answered correctly, the variable value was 1, and 0 otherwise. We then built the variable "knowledge" so that each respondent could obtain a score ranging from 0 to 6. Knowledge was categorized as low if the score was below 2, medium if between 2 and 4, and high if above or equal to 5. The second variable that needed recoding was "social norm." Respondents were asked to indicate how important was the opinion towards fresh-cut products of their relevant people. Specifically, we included questions about family, friends, colleagues, and social media influencers. Following the statistical procedure indicated in Acock (2008), we built the indicator "social norm" using the mean score method. Principal component analysis and alpha Cronbach were determined before calculating the scores. The social norm's eigenvalue was unique with a value of 2.18, and the Alpha Cronbach's between the four items was 0.71. The eigenvalue of knowledge was 1.95, and the correlation between items 0.60.

Table 1 reports the descriptive statistics of the whole sample. There is a slight prevalence of female respondents, and, overall, 57% of participants have less than 30 years old. The presence of young people in the sample is also reflected by the distribution of job seniority level, with only 15% of respondents occupying managerial roles. Income distribution is even among the four ranges of the variable. One-fourth of respondents regularly consume RTE salads. As specified in the question, regular consumption indicated that respondents purchase an RTE salad about three times per week. Consumption is mainly for snacking (60%) or for lunch (60%). Purchasing locations are the supermarket (75%) and convenience stores (25%). About half of respondents regularly practice some physical activity during their free time (48%).

Variable	**Frequency**	**Variable**	**Frequency**
Gender		**Regular consumption of RTE**	
Female	*52%*	salads	
Male	*48%*	*Yes*	*25%*
Age		*No*	*75%*
21-25	*23%*	**Fitness**	
26-30	*34%*	*Yes*	*48%*
31-35	*12%*	*No*	*52%*
36-40	*17%*	**Learning-advertising**	
>40	*14%*	*Yes*	*76%*
Income		*No*	*24%*
Below 15,000rmb	*28%*	**Learning-social media**	
Between 15 and 20,000rmb	*25%*	*Yes*	*31%*
Between 20 and 25,000rmb	*21%*	*No*	*69%*
More than 25,000rmb	*27%*	**Purchasing-supermarket**	
Job seniority level		*Yes*	*75%*
Entry	*40%*	*No*	*25%*
Middle	*45%*	**Purchasing-convenience store**	
Managerial	*15%*	*Yes*	*49%*
Household size		*No*	*51%*
One or two people	*18%*	**Consumption for snacking**	
Three people	*50%*	*Yes*	*60%*
More than three people	*32%*	*No*	*40%*
Knowledge of RTE		**Consumption for lunch**	
Low	*25%*	*Yes*	*60%*
Medium	*39%*	*No*	*40%*
High	*36%*		

		β_j	s.e.
(Intercept)		-3.63^{***}	(0.99)
Gender		-1.04^{**}	(0.39)
Age	26-30	1.32^{*}	(0.58)
	31-35	0.44	(0.64)
	36-40	0.35	(0.58)
	>40	-0.76	(0.72)
Income	15k< Inc. $\leq 20k$	-1.61^{**}	(0.53)
	20k< Inc. $\leq 25k$	-1.12	(0.58)
	Inc. >25k	-0.38	(0.49)
Job Seniority	Middle	1.61^{***}	(0.45)
	Managerial	0.91	(0.64)
Household size	3 people	0.20	(0.57)
	>3 people	1.17^{*}	(0.59)
Knowledge	Medium	0.10	(0.44)
	High	-0.31	(0.50)
Fitness (yes)		1.39^{***}	(0.38)
Learning	Advertising (0-1)	-1.18^{**}	(0.42)
	Social (0-1)	-1.05^{*}	(0.42)
Purchasing	Supermarket (0-1)	1.04	(0.56)
	Conv. store (0-1)	0.51	(0.42)
Consumption	Lunch (0-1)	0.61	(0.38)
	Snacking (0-1)	-0.84^{*}	(0.41)
SN_I2		0.42	(0.70)
SN_I3		1.52^{*}	(0.76)
AIC		284.67	
BIC		372.58	
Log Likelihood		-118.33	
Deviance		236.67	
Num. obs.		288	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Logi model estimation

Table n.2 reports the logit regression results performed on the 70% of the overall sample. The table reports the coefficients (or odds ratio?), standard errors, and significance level. Concerning socio-demographic results, we observed that male respondents are less likely to regularly consume RTE salads compared to female respondents ($p < 0.000$). Respondents between 26 and 30 years old were more likely to consume RTE salads than younger respondents ($p < 0.05$). A lower income was also significantly associated with the consumption of fresh salads. A similar result was observed for respondents in a middle-level job compared to those at the entry-level. Regarding the other variables, we observed that fitness and social norm were positively associated with RTE salad consumption. Social and advertising did not seem to be the main channels through which consumers find information about fresh-cut products. Frequent consumers were more likely to consume RTE salads for lunch and buy them at the supermarket. These results, altogether, are helpful to outline the profile of the typical Chnese consumer of RTE salads. Young, prevalently females, at middle-career level, healthy oriented, pragmatic, and inclined to listen to his/her group reference opinions for food purchasing.

4.2 Prediction

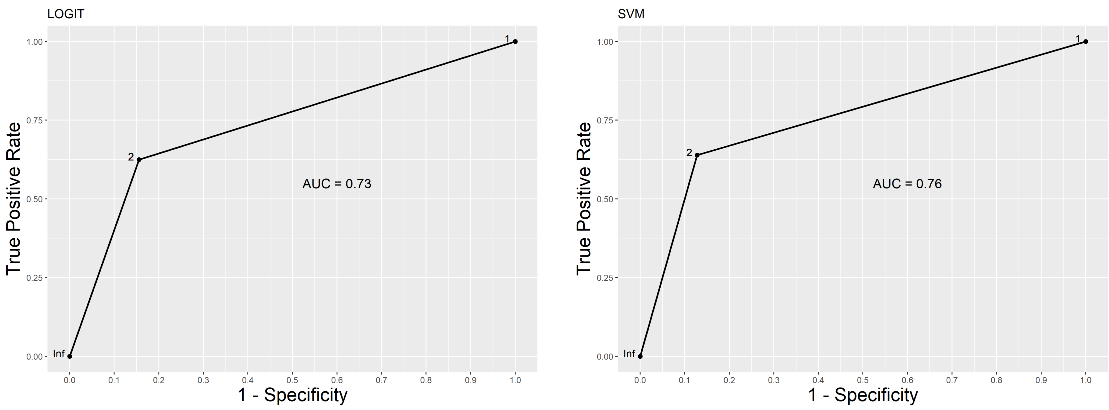

Figure 1: AUC: Logit and SVM

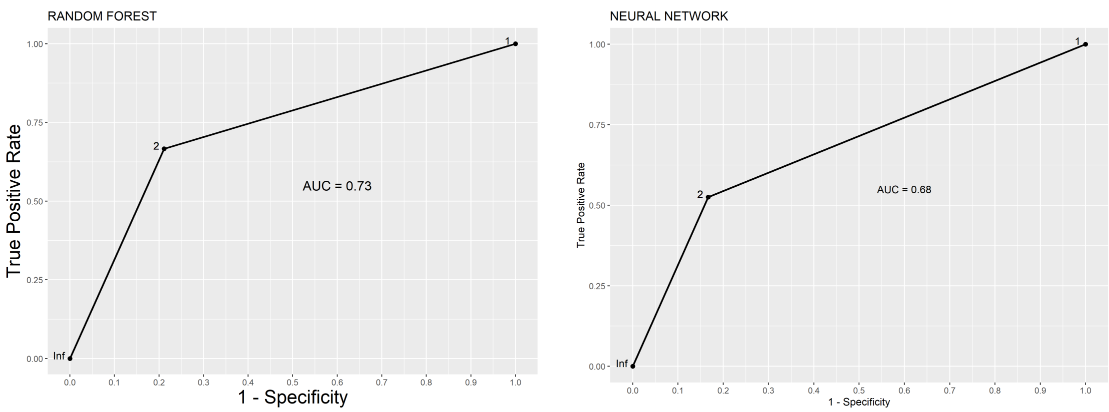

Figure 2: AUC: Random Forest and Deep Neural Network

Model	Accuracy	AUC	Mcnemar's Test P-Value	Sens.-Spec.	Kappa
Logit	0.78 (0.70-0.85)	0.73	0.84452	0.864-0.588	0.4602
Random Forest	0.77 (0.68-0.84)	0.73	0.0045	0.931-0.352	0.4602
SVM	0.80 (0.72-0.87)	0.76	0.83826	0.852-677	0.4602
Neural Network	0.73 (0.64-0.80)	0.67	0.98	0.809-0.528	0.4602

5 Conclusion

6 Appendix

6.1 Support Vector Machine

Support Vector Machines originated as an altertinative to Neural Networks in the wide context of statistical learning, offering a solution to this trade-off by arbitrarily fixing model accuracy and minimising model complexity.
In a classification problem, this translates in an optimal solution identified by a linear hyperplane which correctly classifies observed data and lies as far as possible from them.

Consider the problem of binary classification or dichotomization. Considering the case of a two-dimensional input space, i.e., $\mathrm{x} \in \Re^2$. Training data containing l observation of two explanatory variable X_1, X_2 and the realisation of a target class Y, are given as:

$$\left(\mathrm{x}_1, y_1\right), \left(\mathrm{x}_2, y_2\right), \ldots, \left(\mathrm{x}_l, y_l\right), \mathrm{x} \in \Re^n, y \in \{+1, -1\}$$

Data are linearly separable and we aim at find the optimal separating function (among many different hyperplanes) without knowing the underlying probability distribution $P(\mathrm{x}, y)$.

By using given training examples, during the learning stage, our machine finds parameters $\mathbf{w} = [w_1 w_2 \ldots w_n]^T$ and b of (a discriminant) or decision function $d(\mathrm{x}, \mathrm{w}, b)$ given as

$$d(\mathrm{x}, \mathrm{w}, b) = \mathrm{w}^T \mathrm{x} + b = \sum_{i=1}^{n} w_i x_i + b$$

where $\mathbf{x}, \mathbf{w} \in \Re^n$, and the scalar b is called a bias. After the successful training stage, by using the weights obtained, the learning machine, given unseen pattern x_p, produces output according to the following decision rule:

- if $d\left(\mathrm{x}_p, \mathrm{w}, b\right) > 0$, the pattern x_p belongs to $\{C_0\}$ (i.e. $, o = y_1 = +1$) and
- if, $d\left(\mathrm{x}_p, \mathrm{w}, b\right) < 0$ the pattern x_p belongs to $\{C_1\}$ (i.e., $o = y_2 = -1$).

Where $\{C_0, C_1\}$ are the two possible output classes and b is a bias term.
Under these assumptions we define the setting outlined in Figure 3, where the two dashed lines describe the so-called *margin* to be maximised, which equals to

$$M = \min_{\mathbf{x}:y=1} \frac{\mathbf{wx}}{||\mathbf{w}||} - \max_{\mathbf{x}:y=-1} \frac{\mathbf{wx}}{||\mathbf{w}||}, \tag{7}$$

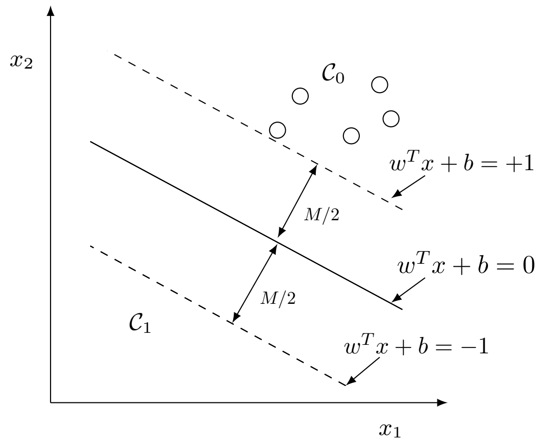

Figure 3: SVM decision boundary and margin.

Since from the decision rule and (7) follows that

$$M = 2\frac{|\mathbf{wx}|}{||\mathbf{w}||} = \frac{2}{||\mathbf{w}||} \tag{8}$$

that is a geometrical translation of margins. Therefore we can write the SVM optimisation problem as

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ sub \quad & y_n[\mathbf{w}^T\mathbf{x}_n + b] \geq 1 \quad \forall n = 1, \dots, N \end{aligned} \tag{9}$$

This is a classic quadratic optimization problem with inequality constraints. Such an optimization problem is solved by the saddle point of the Lagrange functional (Lagrangian):

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{n=1}^{N} \alpha_i \left\{ y_n \left[\mathbf{w}^T\mathbf{x}_i + b \right] - 1 \right\}$$

where the α_i are Lagrange multipliers. The search for an optimal saddle point $(\mathrm{w}_0, b_0, \alpha_0)$ is necessary because Lagrangian L must be minimized with respect to w and b, and has to be maximized with respect to nonnegative α_i (i.e., $\alpha_i \geq 0$ should be found).

6.2 Deep Neural Network

NN training involves an unconstrained optimization problem where the aim is to minimize a function in high dimensional space the so-called loss function, that measures the difference between the predicted values and observed ones. The back-propagation is the most used algorithm for the training of NNs. The algorithm compares the predicted values against the desired ones (objective) and modifies the synaptic weights by back-propagating the gradient of the loss function. Schematically, the procedure alternates forward and backward propagation steps:

- in the forward step, the prediction is computed fixing the synaptic weights,
- in the backward step, the weights are adjusted in order to reduce the error of the network.

The NN iteratively performs forward and backward propagation and modifies the weights to find the combination that minimizes the loss function $\mathcal{L}$.

It is possible to formally describe the forward pass, to aim at obtaining the target output. Specifically, we have the activation function f^k, weights matrix $W^{(k)}$, hidden layers $H^{(k)}$

where $\mathbf{X}$ e $\mathbf{W^{(k)}}$ are respectively, input and weights vectors and $\mathbf{b}$ the bias column. Thus for the k-th hidden layer $\mathbf{H^{(k)}}$:

$$\mathbf{H^{(k)}} = f^k(\mathbf{W^k}\,\mathbf{H^{(k-1)}} + \mathbf{b^{(k)}}) \tag{10}$$

and thus, we can rewrite as function of the input $\mathbf{X}$:

$$\mathbf{H^{(k)}} = f^{(k)}(\mathbf{W^{(k)}}\ \underbrace{f^{(k-1)}(\dots f^{(1)}(\mathbf{W^1}\,\mathbf{X} + \mathbf{b^{(1)}})\dots)}_{=\mathbf{H^{(k-1)}}} + \mathbf{b^{(k)}}) \tag{11}$$

After the estimation procedure, which implies to estimate the weights, we obtain the final output, $\hat{y}$, resulting from the application of the DNN parameters (weights matrix $\hat{\mathbf{W}}$ and bias $\hat{\mathbf{b}}$) obtained by the optimization procedure described in the following.
NN, like other machine learning techniques, requires the splitting of the dataset into a training and a testing set. The training set stands for supervised learning, while the testing set is used to validate the model. After the training phase, the network has learned the input–output functional relationship and it should be able to predict future values using only the input. Practically, NN parameters are obtained by the minimization of the overall loss function $\mathcal{L}$. In the classification problem the most widely used loss function is the binary cross-entropy:

$$\mathcal{L}(y, \hat{y}) = -\sum_{i=1}^{n} y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$$

To minimize the loss function, we use the Gradient Descent optimization algorithm, which iteratively moves in the direction of steepest descent as defined by the negative of the gradient.

This algorithm proceeds by minimizing $\mathcal{L}$ at each step t, therefore differentiating the loss function with respect to the weights ($\mathbf{W}$). For the generic weights $w_{n,n}^{(k)}$ and the k-th layer, the algorithm proceeds using the chain derivation rule described in the following equation:

$$\frac{\partial \mathcal{L}(y, \hat{y})}{\partial w_{n,n}^{(k)}} = \frac{\partial \mathcal{L}(y, \hat{y})}{\partial H_n^{(k)}} \frac{\partial H_n^{(k)}}{\partial z_n^{(k)}} \frac{\partial z_n^{(k)}}{\partial w_{n,n}^{(k)}} \tag{12}$$

where $z_n^{(k)} = w_n^{(k)}\,H_n^{(k-1)} + b_n^{(k)}$. To update the weights ($\tilde{\mathbf{W}}$), the gradient of the loss function, $\nabla\mathcal{L}_t(y, \hat{y})$, is multiplied by a scalar, η, often called learning rate, according to the following scheme:

$$\tilde{\mathbf{W}} = \mathbf{W} - \eta\nabla\mathcal{L}_t(y, \hat{y}) \tag{13}$$

In a figurative way, the idea behind the gradient descent is similar to "climbing down a hill" until a global or local minimum is reached. At each update, the search moves in the opposite direction of the gradient and the learning rate η determines the amplitude of this movement, controlling the adjustment in the weights, thus determining how fast or slow we will move towards the optimal weights. A very large learning rate leads to a sub-optimal solution. A very small learning rate involves too many iterations to find the optimal solution. Then, the learning rate can be considered the most important hyper-parameter for tuning NN.

The search for the optimal parameters is then carried out through an optimization process where the NN initial weights are selected in an arbitrary (random) way so they are not optimal parameters. The iterations of the algorithm lead to the optimization of the weights and minimization of the error. The choices concerning the type of architecture (e.g., the number of hidden layers, units for each layer) and the hyperparameter (e.g., learning rate, activation functions, and loss function), remains a heuristic problem for NN users: the choice often depends on the type of data and it is not an easy step.

6.3 Random Forest

The random forest (RF) algorithm creates a collection of decision trees from a casually variant of the tree. Once one specific learning set is defined, the RF presents a random perturbation to the learning procedure and in this way a differentiation among the trees is produced. Successively the predictions of all these trees is derived through the impelementation of aggregation techniques. The first aggregation procedure was described by Breiman (1996); the authors proposed the well know bagging based on random bootstrap copies of the original data to assemble different trees. Later in 2001 the same authors Breiman (2001) proposed the random forest as an extention of the procedure of the bagging such that it combines the bootstrap with randomization of the input variables to separate internal nodes t. This means that the algorithm does not identify the best split $s_t = s^*$ among all variables, but firstly creates a random subset of K variables for each node and among them determines the best split.

The RF estimator, for both regression and classification problem, of the target variable $\hat{y}_{R_j}$ is a function of the regression or classification tree estimator, $f^{\text{tree}}(\mathbf{X}) = \sum_{j \in J} \hat{y}_{R_j} \mathbf{1}_{\{\mathbf{X} \in R_j\}}$, where $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ is the vector of the predictors, $\mathbf{1}_{\{-\}}$ represents the indicator function and $(R_j)_{j \in J}$ are the regions of the predictors space obtained by minimizing the binary cross entropy Loss Function Therefore, denoting the number of bootstrap samples by B and the decision tree estimator developed on the sample $b \in B$ by $\hat{f}^{\text{tree}}(\mathbf{X} \mid b)$, the RF estimator is defined as follows:

$$f^{RF}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} f^{\text{tree}}(\mathbf{X} \mid b)$$

References

Artés, F., Gómez, P., Aguayo, E., Escalona, V., Artés-Hernández, F. (2009). Sustainable sanitation techniques for keeping quality and safety of fresh-cut plant commodities. Postharvest Biology and Technology, 51(3), 287–296.

The Ministry of Health of China. Management Measures for Nutrition Improvement. The Ministry of Health of China, Beijing;2010.

The State Council of China. National Nutrition Plan (2017–2030). The State Council of China; 2017.

The State Council of China. China Nutrition Improvement ActionPlan (1996-2000). The State Council of China, Beijing; 1997.

The State Council of China. Healthy China Action Plan (2019–2030). The State Council of China, Beijing; 2019.

World Health Organization (2008). WHO European Action Plan for Food and Nutrition 2007/2012.