ISCTE Business School
Instituto Universitário de Lisboa

UNVEILING THE FEATURES OF SUCCESSFUL EBAY SELLERS OF SMARTPHONES – A DATA MINING SALES PREDICTIVE MODEL

Ana Teresa Nunes Biscaia Correia da Silva

Dissertation submitted as partial requirement for the conferral of

Master in International Management

Supervisor:
Paulo Rita, Full Professor, ISCTE Business School, Department of Marketing, Operations and General Management

Co-supervisor:
Sérgio Moro, Assistant Professor, ISCTE – IUL, Department of Information Science and Technology

September 2016

ABSTRACT

EBay is one of the largest online retailing corporations worldwide, providing numerous ways for customer feedback on registered sellers. In accordance, with the advent of Web 2.0 and online shopping, an immensity of data is collected from manifold devices. This data is often unstructured, which inevitably asks for some form of further treatment that allows classification, discovery of patterns and trends or prediction of outcomes. That treatment implies the usage of increasingly complex and combined statistical tools as the size of datasets builds up. Nowadays, datasets may extend to several exabytes, which can be transformed into knowledge using adequate methods. The aim of the present study is to evaluate and analyse which and in what way seller and product attributes such as feedback ratings and price influence sales of smartphones on eBay using data mining framework and techniques. The methods used include SVM algorithms for modelling the sales of smartphones by eBay sellers combined with 10-fold cross-validation scheme which ensured model robustness and employment of metrics MAE, RAE and NMAE for the sake of gauging prediction accuracy followed by sensitivity analysis in order to assess the influence of individual features on sales. The methods were considered effective for both modelling evaluation and knowledge extraction reaching positive results although with some discrepancies between different prediction accuracy metrics. Lastly, it was discovered that the number of items in auction, average price and the variety of products available from a given seller were the most significant attributes, i.e., the largest contributors for sales.

Keywords: Online Sales; EBay Sellers; Data Mining; Smartphones; Marketing[1]***, Data Analysis***[2]

[1] Keywords according to JEL Classification guidelines (M310).
[2] Keywords according to JEL Classification guidelines (C380).

RESUMO

O EBay é uma das plataformas e retalho online de maior dimensão e abarca inúmeras oportunidades de extração de dados de feedback dos consumidores sobre vários vendedores. Em concordância, o advento da Web 2.0 e das compras online está fortemente associado à geração de dados em abundância e à possibilidade da sua respetiva recolha através de variados dispositivos e plataformas. Estes dados encontram-se, frequentemente, desestruturados o que inevitavelmente revela a necessidade da sua normalização e tratamento mais aprofundado de modo a possibilitar tarefas de classificação, descoberta de padrões e tendências ou de previsão. A complexidade dos métodos estatísticos aplicados para executar essas tarefas aumenta ao mesmo tempo que a dimensão das bases de dados. Atualmente, existem bases de dados que atingem vários exabytes e que se constituem como oportunidades para extração de conhecimento dado que métodos apropriados e particularizados sejam utilizados. Pretende-se, então, com o presente estudo quantificar e analisar quais e de que modo as características de vendedores e produtos influenciam as vendas de smartphones no eBay, recorrendo ao enquadramento conceptual e técnicas de mineração de dados. Os métodos utilizados incluem máquinas de vetores de suporte (SVMs) visando a modelação das vendas de smartphones por vendedores do eBay em combinação com validação cruzada 10-fold de modo a assegurar a robustez do modelo e com recurso às métricas de avaliação de desempenho erro absoluto médio (MAE), erro absoluto relativo (RAE) e erro absoluto médio normalizado (NMAE) para garantir a precisão do modelo preditivo. Seguidamente, é implementada a análise de sensibilidade para aferir a contribuição individual de cada atributo para as vendas. Os métodos são considerados eficazes tanto na avaliação do modelo como na extração de conhecimento visto que viabilizam resultados positivos ainda que sejam verificadas discrepâncias entre as estimativas para diferentes métricas de desempenho. Finalmente, foi possível descobrir que número de itens em leilão, o preço médio e a variedade de produtos disponibilizada por cada vendedor foram os atributos mais significantes, i.e., os que mais contribuíram para as vendas.

Palavras-chave: Vendas Online; Vendedores no EBay; Mineração de Dados; Smartphones; Marketing[1]***; Data Analysis***[2]

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisors, whose continuous support and motivation were fundamental to develop my knowledge in this particular field and subsequently bring it to fruition through a successful completion of this study and the respective stage of my academic path.

INDEX

INDEX OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

MAE	Mean absolute error
MAPE	Mean absolute percentage error
NMAE	Normalized mean absolute error
RAE	Relative absolute error
RBF	Radial basis function
SA	Sensitivity analysis
SVM	Support vector machine
eWOM	Electronic word of mouth

INTRODUCTION

1.1 Overview of components Web 2.0, online shopping and smartphones

With the advent of Web 2.0 and online shopping, an immensity of data is collected from myriad applications and devices. EBay is an excellent example of an online company boosting its way through the Web 2.0 era, being currently one of the largest online sales platforms, supplying online retailing services for any seller worldwide (Einav et al., 2014). Such a colossal player allows a large number of different means for users to contribute with feedback on the services provided and registered sellers.

Technologies and telecommunications have become essential elements of everyday life and business. The need for increasingly fast and optimised devices has guaranteed a steady growth in the technological industry, although at due different regional paces (Kellerman, 2010). Mobile devices have also become one of the primary sources for online shopping (Pearce & Rice, 2013). In UK, for example, mobile has already surpassed desktop by 44% (The Guardian, 2014). Such relevance can prove to be an effective driver for increasing sales of mobile devices (Bilgihan et al., 2016). Smartphones belong to this category since they are essentially *"*mobile phones with more advanced computing capabilities and connectivity than regular mobile phones*"* (Statista, 2016). They have been available in consumer markets since the 1990s nevertheless, only became truly popular and mainstream when, in 2007, the iPhone's introduction by Apple transformed the industry, leading to the first Android based smartphone being released to consumer markets in late 2008 (Lee et al., 2015).

It is estimated that by 2017 a third of the population worldwide will own a smartphone, which will, according to forecasts, encompass 2.6 billion smartphones (Statista, 2016). In 2015, solely, the global smartphone industry was responsible for the generation of approximately 240.55 billion Euros although with a decrease by roughly 1.49% comparatively to the previous year.

Online e-commerce platforms such as Amazon, Taobao and eBay are references in online shopping and auctions for several products including smartphones, containing and producing substantial amount of data, which may be used for knowledge extraction using appropriate data mining techniques.

1.2 The need for data mining

Data mining is the process of discovering patterns of knowledge from raw data (Sharda et al., 2014). Its roots lie on statistics and data analysis, and have been greatly enhanced through machine learning techniques and methods. Data mining as an evolving process has been around for some time, but only since the 1990s, when the concept was coined, until today has it been gaining considerably more popularity and attention (Fayyad et al., 1996; Sharda et al., 2014). This is happening due to the large amounts of data (in what is known as big data) that are generated every second from several sources, such as manifold sensors and devices (Pal et al., 2014) and also social media and smartphones' applications (Chen et al., 2012).

In sum, "the world is data rich but information poor" (Han et al., 2012). This idea is the stepping stone for this data mining project and, therefore, its goal is to generate the type of information that is able to leverage decision-making through actionable knowledge, which might be of particular interest for online retail sellers, online marketplaces and marketing practitioners, who may use the insights provided in the process of discovering how online features of sellers influence sales, which will be at the core of the analysis. In fact, large online e-commerce websites represent the future of retailers (Tadelis, 2016), and top players such as eBay, Amazon and Alibaba are among the most technologically innovative organisations worldwide (Liu & Lu, 2015), given its technological nature. Therefore, research on improving customer service based on cutting edge technology can help cope with the challenges of tomorrow.

Traditional data mining projects are time-consuming as all the data is most of the times manually extracted and with limited amount of resources, which usually leads to limitations in the scope of analysis. In this case, the research is narrowed to the extraction of knowledge in the form of features' relevance from sellers of smartphones on eBay, one of the largest e-tailers worldwide. EBay yielded a market share on auctions of more than 99% in the US in 2008, being also one of the key players in online retail, according to Haucap and Heimeshoff (2014).

1.3 Objectives and contribution

The aim of this study is to provide insights about what it takes to be a successful eBay seller by unveiling through data mining which seller attributes contribute the most to actual sales, i.e. which have the most influence on the number of items sold. Previous literature has approached the subject mostly from consumer and customers' perspectives yet rarely from the sellers' point of view. As the number of registered sellers on online platforms rises worldwide, it becomes crucial to understand what drives the success of sellers within the different dimensions that can influence their results (Ye et al., 2013; Chen et al., 2014). Such knowledge can be valuable both from a seller's perspective as well as for managing online platforms as for instance offering premium services to the most prospective sellers or improving feedback services and information supplied to the registered users.

There is research focused on consumer demographics in online shopping (Black, 2007), confluence of retailer characteristics, market characteristic and online pricing strategies (Venkatesan et al., 2006), typology of complaints on auction websites (Gregg et al., 2008), cross-cultural transactional behaviours on eBay (Yan et al., 2009) and, most relatedly, prediction of online sales based on reviews in the movies domain (Yu et al., 2010). Yet a stream of research that grasps onto the conspicuous and measurable characteristics of online sellers combined with product attributes in order to determine their impact on sales using data mining predictive techniques is scarce in the literature. Therefore, the immediate purpose of this is to fill in that research gap. In addition, the contributions for the literature are the following:

- Extraction of online seller attributes from an online sales renowned platform, eBay;
- Evaluation of the smartphones' online market through the analysis of eBay sellers' performance.

The next section dives deeply into the theoretical background, which supports the relevance of the subject along with the data mining techniques in use and is followed by a detailed description of the chosen methodology and approach. Then, the results are discussed and interpreted in order to extract adequate knowledge out of the data. Finally, the conclusions are drawn in the last section.

THEORETICAL BACKGROUND

2.1 Web 2.0 and online shopping

Web 2.0 is defined as a "set of applications and technologies that allows users to create, edit, and distribute content; share preferences, bookmarks, and online personas; participate in virtual lives; and build online communities" (Laudon et al., 2007). It is considered a new stage of development of the web and it differs from the previous one by the drastic increase in information density, interactivity and level of customization. This new phase can be traced back to 2007, when the changes became evident. Hence, the relevance of drawing attention to the associated shift from making online purchases to going shopping online (Hemp, 2006) as the online environment and virtual communities become vital elements in the consumer journey. Thus recommendations from other consumers, instead of advice from family and friends, are also becoming an increasingly important decision factor (Kotler et al., 2012). It is important to examine the e-tail environment as a whole since "electronic markets enable volumes and speeds that human middlemen could not accomplish" (Venkatesan et al., 2006; Hess et al., 1994). However, there is still plenty of research focused exclusively on brick and mortar retail context when compared with pure online play and bricks-and-clicks, which have been growing expressively in the last years (Grewal et al., 2010).

The steadily growing number of internet users and dispersion of mobile devices fuels the aligned rise of sales in retail e-commerce. It is pointed out that internet has become more frequently used and with increased convenience since computers and smartphones have become more accessible alongside the worldwide modernization of countries (Statista, 2016). As Web 2.0 has been facilitating the widespread dissemination and acceptance of online shopping, the latter has been around for quite a longer time, since the 1980s to be precise although the technology is available since ca 1979 (Aldrich, 2011). The Web is a unique form of social space within the cyberspace that embodies both resource and production forces and, subsequently, enables online shopping which is part of that same space (Kellerman, 2010).

In 1994, Amazon was founded, followed by eBay, AuctionWeb at the time, in 1995. Since then the platforms have evolved tremendously in a way that changed people's shopping habits. Currently, out of the 13 largest online retail and auction sites, eBay is ranked 8th

with a global market reach of 13.6% (Statista, 2016). In addition, and according to Alexa Internet's rank, which evaluates internet traffic, in March 2016, Amazon.com was the 6th, Taobao.com the 12nd and EBay.com the 24th out of the 500 top sites on the web worldwide. This undeniably shows the significance of online shopping in overall web traffic scene and reinforces the idea that online shopping is an extremely relevant topic with unique features and applications, which are in need of a thorough closer look.

Looking from the consumers' perspective, Cheung et al. (2005) pointed out that the main determinants of online consumer behaviour were related with consumer characteristics, environmental influences, product/service characteristics, medium characteristics and merchants and intermediates' characteristics, which would have a transversal impact through the online customer journey.

2.1.1 Online auctions, reputation and pricing

Wilcox (2000) highlighted that auctions had been gaining a level of popularity consistent with an upward growth in the market for consumer goods due to the dispersion of internet-based auctions. The English auction type is the most frequently used but different forms of auction dynamic-enabled pricing are expected to emerge (Chen, 2002). EBay, for example, is a combination of English and Vickrey auction types. Online auctions offer the possibility of cost reduction while simultaneously increasing the number of potential bidders (Jayaraman et al., 2003). Subsequently, sources of competitive advantage for retailers in the marketplace are being redefined as traditional approaches become challenged (IBM, 2016) and crafting the online shopping experience becomes imperative (Zhang et al., 2010). This involves the conception of virtually adaptable marketing strategies with regards to the mix variables, namely product presentation (Park et al., 2005) and website/app design along with value-adding features such as technical support.

In online auction marketplaces, reputation systems are used as communication means of performance to potential customers (Huang et al., 2010). According to Ratchford (2009) online auctions have been approached from the perspectives of economics (Bajari et al., 2004), consumer behaviour (Cheema et al., 2005) along with management science (Pinker et al., 2003) and he adds that they are usually employed when supply is insufficient and the number of potential buyers is small and widely disseminated. The same author draws attention to the acute importance of reputation in auctions due to the associated risk of

moral hazard and suggests that an appropriate measure to deal with the issue is enablement of seller ratings. However, buyers also face the winner's curse risk resultant of overvaluing an item.

Opposite to brick and mortar stores, in e-tail, reputation and trust cannot be as easily established since quality cannot be assessed through the same type of cues (Bruce et al., 2004). It was also stressed that their establishment might possibly be of higher importance in the latter. Thus, e-tail requires the usage of a different set of mechanisms highly based on customer feedback including ratings and reviews. Those can only be obtained if there is an underlying marketing strategy emphasizing price and non-price attributes. The latter is linked to quality of items and delivery as focal points for customers and the former is associated with reassurance of price competitiveness (Bruce et al., 2004). On that account, firm, product and channel factors are identified as backbones for the development of retail pricing strategies (Grewal et al., 2010). Consequently, differentiation can be achieved through information available, price, assortment, convenience and experience, which typify the retail mix. The cumulative experience that sellers gain from the process of crafting the virtual shopping experience is also an essential part in building reputation.

In early research about pricing it was often argued that the advent of internet would lead to heightened competition online, which would induce price reductions (Brynjolfsson et al., 2006). Nevertheless, it was found that other features of online markets had more impact in the buying decision process such as variety and convenience. Hence, the trade-off between breadth and depth could stand a chance at being solved (Grewal, 2010). The turn up of Web 2.0 tools that accelerate information sharing and networking has been affirmative in self-generation of content from both sellers and customers, which is the foundation for creating the set of mechanisms that enables building online reputation.

Moreover, recent studies have devoted efforts in finding influencing features on the prices of online sales. Kocas and Akkan (2016) evaluated how online feedback and rating from customers affected the prices of books from twenty-four categories sold through Amazon.com. Their work has proved that customer ratings should be accounted for in order to increase profitability. Cao et al. (2015) presented a study on dynamic pricing of online shopping by dividing customers in patient and inpatient potential buyers, providing evidences that the optimal pricing policy should limit dynamic pricing in cases of customers with less patience. Sellers' reputation has proven to be an effective influencer

of the pricing policy followed, with highly reputed sellers having advantages in pricing, as shown by Xu et al. (2015) through an analysis of Taobao sellers. However, the same study also emphasizes that literature on pricing relation to reputation is scarce. Previous article published by Ye et al. (2013) has reached a similar finding by analysing both Taobao and eBay sellers. Both works are conclusive in that sellers' features do affect pricing, influencing sales performance, with the latter adopting a regression model for studying three sellers' attributes: reputation score, number of positive reviews, and score for "item as described". However, this study did not consider further attributes from sellers that are available on eBay, such as the neutral and negative reviews. Furthermore, both studies analysed online sellers in a pricing perspective, not accounting for the number of sales derived from sellers' features.

Chong et al. (2015), have used big data from Amazon.com to evaluate the impact of online promotion marketing along with online reviews on product demands and they have discovered that reviews played a crucial role surpassing online promotional marketing. They also reinforced the idea that user-generated content is gaining more and more popularity among consumers as the addition of user reviews together with detailed descriptions and price information are becoming standard. The ease in accessibility to increasingly rich and diverse sources of information, i.e., electronic word of mouth (eWOM) is contributing substantially for optimising the customer journey as consumers are finally finding the information they need in a much faster way. On the other hand, handling brand health and reputation in this uncertain marketing environment is one of the main challenges for sellers (Leeflang et al., 2014). Previous literature (Chong et al., 2015; Lu et al., 2013; Zhu et al., 2010; Chevalier et al., 2006) has proven the linkage between reviews and ratings on product sales. Chen et al. (2004) show that quantity of recommendations is positively linked with sales of books on Amazon.com. However, Chong et al. (2015) point out that the role of eWOM and its attributes in search products such as electronic products has been neglected in research when compared with experience products such as books and movies, which have been earning significantly more attention. Volume of reviews is also mentioned as an important cue for consumers, addressing the unequivocal relationship between the amount of discussion about a product or service and the level of awareness raised, which would consequently lead to visible changes in sales.

The experience of sellers is another critical aspect in electronic markets since the performance of new sellers on eBay is typically poorer than that of experienced ones (Goes et al., 2013). Additionally, Huang et al. (2010) revealed that fulfilment speed, reliability and communication had different impacts on satisfaction and dissatisfaction of customers, using two-factor theory as a framework. Speed was found to be an important satisfier while reliability and communication were identified as critical dissatisfiers. However, their approach lacked insight into particular product categories, which leaves room for further analysis.

2.1.2 The importance of assortment management

Managing the variety of products sold, i.e., the assortment has always been an essential element of business development (Ramdas, 2003). In today's environment where high levels of demand together with increased want for a personalized offer have become the norm, finding the right balance between variety and the level of customization is often a challenge. High variety can be associated with increased variability and lead to errors in forecasting (Ramdas, 2003; Fisher, 1997). Therefore, firms must assess how much revenues outweigh the costs when choosing the assortment strategy that best fits the established goals.

The adoption of niche versus mass strategies is another important aspect related with assortment management and it is inextricably linked with the level of variety and specialization of the products sold. Furthermore, and adducing preceding paragraphs, it was discovered that, from the demand side, huge variety of inefficiently organised items can stagger consumers and hold back purchases due to forecasting errors and difficulty for consumers to find the products they are looking for (Brynjolfsson et al., 2006). If sellers choose marketing and assortment strategies that are not compatible in ensuring a smooth supply chain (Fisher, 1997), it can have a negative impact on consumer behaviour and repurchase intention based on satisfaction (Yen et al., 2007) which will inevitably affect sales, e.g., when the online seller has several product categories for which he can't ensure a smooth shipping due to factors associated with shipping costs, local geography or transportation and he could instead increase variety but specialise in selling particular products in localised markets for which he can ensure it without compromising customer satisfaction. This is particularly relevant in the case of sellers with lower level of experience in e-commerce platforms such as eBay. Therefore, creating ingenious active

and passive search tools, in alignment with a healthy balance between mix variables, assortment selection (Zhang et al., 2010) and with the ability to engage and compel consumers in a way that translates into sales, is of pressing importance.

2.2 Smartphones' relevance

The purpose of smartphones' usage has been expanding progressively beyond its core as they become one of the most popular mediums for purchasing products or services, social media activity or conducting research (Bilgihan et al., 2016) with levels of ubiquitous connectivity and convenience never experienced before (IBM, 2016). On top of that, there is the fact that the price of smartphones has been steadily decreasing (Statista, 2016) along with the natural increase in worldwide usage. Regarding sales level, China led the regional race with an increase from 90.1 billion USD in 2013 to 117.8 billion in 2016, followed by North America, Western Europe and Middle East & Africa in the same period. Central & Eastern Europe together with Latin America yielded the lowest sales level.

Between 2012 and 2015, Samsung has been the market leader with regards to worldwide shipment of smartphones followed up Apple. However, the shares have been decreasing has new players such as Xiaomi enter the market. According to retail sales volume from 2010 to 2015, the most expressive brands were Samsung followed by Nokia/Microsoft, which has been gradually decreasing during that period. Apple registers comparably lower volumes yet higher revenues as the second largest global seller of smartphones with over 230 million iPhones sold worldwide in 2015 (Statista, 2016). Brands such as LG, ZTE, Huawei, Sony/Sony Ericsson, Lenovo Motorola, HTC and Xiaomi were also quite relevant concerning volumes. Then, there are regionally relevant players such as Micromax in India and Yulong in China.

Within the 19 most popular online shopping categories, IT and mobile is ranked 5th achieving 40% in global online purchase rate, which reveals the potential and relevance of the category for online shopping. Smartphones are mobile phones with operating systems similar to PCs and they are, therefore, included in the mentioned category. Their number of sales has been increasing sharply as in 2013 it already doubled compared to 2011. It is expected that by 2017 the market penetration of the devices will be of 65.8% in Europe and 62.2% in North America. Thus, it is clear that smartphones are gaining

more and more popularity and, as such, it is foreseeable that their sales will increase and that gathering valuable market information will be a source of added value in marketing planning. Furthermore, it is important to mention that over 335 exabytes of data are generated and stored on a yearly basis through smartphones only (Poelker, 2013). This immensity of data is transversal to all industries and can be extremely valuable if used to retrieve important information (Pal, 2014).

2.3 Data mining techniques and metrics

According to Yu et al. (2010), online reviews and feedback have embedded in them the unique opportunity of extracting business intelligence. The role of data mining becomes evident when wanting to derive that information in order to generate "actionable knowledge" that can be easily accessed and handled by decision makers. Thus, it becomes clear that "data mining is the core of business intelligence" (Han et al., 2012).

Data mining potential extends to pretty much any scientific and business area. From astronomy to marketing, fraud detection, manufacturing or telecommunications (Fayyad et al., 1996), its usefulness transcends any field one might contemplate. Within business applicability: increasing customer intelligence, improvement of operational efficiencies and customer customization are only some of the broad possibilities for data mining (Pal et al., 2014).

Several examples are known of research taking advantage of data mining, such as for modelling user rating profiles (Marlin, 2004), designing of products and information systems (Kusiak & Smith, 2007), predicting wind power (Louka et al., 2008), predicting bank telemarketing successful contacts (Moro et al., 2014) or measuring social media performance (Moro et al., 2016). Other examples include the application to e-learning domain (Hanna, 2004), customer knowledge creation (Khodakarami & Chan, 2014) or for discovering the helpfulness of online reviews (Lee & Choeh, 2014). These are only a few among a vast array of studies in which data mining was used.

In sum, data mining and its techniques can be applied to any science and any industry as there is still a plethora of untapped opportunities. However, the particular application should always be taken into account since there is arguably a universal data mining method so far. Therefore, selecting the most suitable one can be considered somewhat of an art (Fayyad et al., 1996).

In order to perform the inherent data mining tasks, numerous methods can be used. Such procedures usually entail machine learning algorithms, which resort to computational methods that allow learning information straight from the data without the need to have a pre-set equation serving as a model (Mathworks, 2016). This enables improvement of performance as more and more observations are added to the dataset, allowing the machine to learn. Decision trees, neural networks and support vector machines are just a few of them. In the expanse of this project, only SVMs and sensitivity analysis will be explained in detail. Within these methods there are several functionalities to handle the patterns which are found throughout data mining tasks. Those functionalities are fundamentally categorized into descriptive and predictive. The first ones are associated with the description of properties of the data in a target dataset while the second "ones use induction on the current data in order to make predictions" (Han et al, 2012).

Essentially, data mining is a vital element in the knowledge discovery process, which allows data to grow into intelligible information. It is like turning a set of ingredients into a satisfying meal using the most adequate recipe(s) making the most with limitations of those ingredients and their relationships to given purpose(s). Data mining is, in a way, "cooking" your data.

2.3.1 Support vector machines

Vapnik and Cortes (1995) are the "architects" behind the support-vector network learning machine in pivotal stages of SVMs. They presented the idea of mapping nonlinear input vectors into a high-dimension feature space where a linear decision surface would be built within a deeply widening scenario in which training data could be separated with errors, and therefore breaking ground to solving real problems, inspired by the initial discoveries of Fisher (1936) for pattern recognition algorithms.

*"*Support Vector Learning Machines (SVM) are finding application in pattern recognition, regression estimation, and operator inversion for ill-posed problems" (Schölkopf et al., 1997), which was an early sign of their increasing popularity and applicability. They are often used for classification of linear and nonlinear data through the transformation of the original data into a higher dimension using a kernel function that computes dot products in the transformed space (Friedman et al., 2001), from where it can find a hyperplane for data separation using essential training tuples called support

vectors (Han et al., 2012). They can also be used for regression with the requirement of adding a loss function (Smola et al., 1997).

A support vector machine is an algorithm, which belongs to the same typology as other neural network classifiers, e.g., an SVM with a Gaussian radial-basis function (RBF) displays a matching hyperplane to the neural network identified as RBF network. (Han et al., 2012). The completeness of SVMs' algorithms enables the construction of models with enough complexity that, are, however, simple in a way which makes mathematical analysis possible. The algorithm comprehends a significant amount of neural nets, RBF net and also polynomial classifiers. (Hearst et al., 1998).

One of the main benefits of SVMs is that they can attain good performance levels when applied to real problems just as they can be analysed with higher complexity and employing theoretical concepts from computational learning theory (Hearst et al., 1998). The idea is often supported as SVMs are considered as an attractive approach to data modelling. They combine generalisation control with a technique to address the curse of dimensionality (Gunn, 1998). However, more recent research has confined the statements to questionability since it is argued that in the presence of powers and products by giving the same weights to terms in 2XjX′j form, a polynomial kernel of degree 2 in a 2-input feature space won't be able to adapt to subspace concentrations and will difficultly find structure by having many dimensions where to search (Friedman et al., 2001). Knowledge would have to be assembled into the kernel to solve the problem of multidimensionality. On the other hand, the same author backs the idea that, at the same time, SVMs performed well when applied to real learning problems.

In essence, an SVM encompasses a set of techniques that allow building a linear boundary in a large transformed version of the feature space in order to produce nonlinear boundaries using a kernel (Friedman et al., 2001) and this simplifies analysis because it enables the correspondence of a linear method to a high-dimensional feature space nonlinearly related to the input space (Hearst et al., 1998). In other words, what occurs is that through the usage of a kernel, a similarity or approximation function, and addition of a loss function, the hinge loss function, the desired outcome is optimised. The jumbled data in the input space is transformed into nonlinear boundaries separated by an optimal hyperplane, the feature space, which will be easier to analyse because the data becomes

structured and, therefore, further analysis is made possible until it develops into intelligible information.

2.3.2 Sensitivity analysis

When dealing with black box models, it is often a challenge to extract knowledge in a way that is easy to understand. That fact inspired a new stream of research to tackle the inherent problem. As a consequence, methods such as extracting rules from networks and sensitivity analysis (SA) have emerged.

Sensitivity analysis enables the assessment of the importance of input factors to a given model (Saltelli et al., 2000) and also their effects on the model's responses (Cortez et al., 2011). It is frequently employed in order to evaluate the coherence and attractiveness of a kernel-based and ensemble black box models such as SVMs or neural networks and, subsequently, facilitating their interpretability. This is a central element to any model since it will contribute for increased comprehension by different audiences and trust in data mining. It can be disclosed using one among sensitivity analysis algorithms together with appropriate visualization techniques. It is pointed out that extraction of rules is rather simplistic and might fail at assessing the representativeness of the model due to disregard of relevant rules and danger of generalisation mainly resultant of discretisation of the separating hyperplanes. (Cortez et al., 2013).

One of the advantages of SA is its broad applicability to almost any supervised learning model as the relationship goes straight to the bottom line of input-output relationships, i.e., the way that any variation in a given input changes the respective output. There are several types of methods, i.e., algorithms to choose from within SA. Among that group are included one-dimensional sensitivity analysis (1D-SA), global sensitivity analysis (GSA) and data-based sensitivity analysis (DSA). They diverge in their suitability to different goals. 1D-SA is very fast but cannot measure complex interactions among the features, whereas GSA is the perfect SA method in terms of interaction measurement but it is computationally too costly. DSA is similar to 1D-SA but it uses training samples in detriment of a baseline vector. In effect, the main goal in the case of DSA is to harvest the possible interactions between inputs but in a faster manner than with GSA. If needed, DSA can even be speeded up if a proportion of the training samples (randomly selected) are used instead of the whole training set. Such feature makes DSA computationally much

more efficient than GSA, while having a better performance than 1D-SA due to its capability of detecting input variable interactions.

All things considered, one can assuredly state that regardless of the SA method in use, they all contribute for higher effectiveness in interpretability of complex models as they all strive at obtaining a set of sensitivity responses.

2.3.3 Regression performance metrics

One of the crucial steps in model building is assessing its adequacy in predicting what it is supposed to. Therefore, one can state that performance and adequacy in prediction models are inevitably connected, i.e., if one model fails to predict its output then it is inadequate (Diebold et al., 2012). This brings out the importance of forecast accuracy since the derived forecasts are used for guiding decision-making.

Although there are plenty of performance measures (Hyndman et al., 2006) error measures are quite often chosen. Multivariate error distributions, which are produced by any forecasting method, enable this process. This measures were created to assess the discrepancies between predicted and actual values. Additionally, they are relevant in model calibration and refining. As a consequence, choosing the most appropriate measures for forecasting accuracy is critical (Armstrong et al., 1992).

There are also performance measures that contribute for model validity and subsequently, its overall accuracy, such as cross-validation. Models are frequently evaluated using this method and its estimates of a prediction error (Fayyad et al., 1996; Weiss et al., 1991). According to Refaeilzadeh et al. (2009), cross-validation can be applied to estimate performance, model selection, and tuning learning model parameters. Moreover, it is also considered a reasonable technique to deal with overfitting. With *k*-fold cross-validation, all the observations are randomly split into *k* equal sets, which are used for training and testing. The latter is used to assess model reaction to new data, constituting a realistic predictive testing approach (Berry et al., 2004). The remaining subsets are used as training data, which is used for model building. Essentially, each of the subsets will operate as both training and testing data but only once as the latter. The equally sized *k* testing subsets are gathered generating an estimate for the whole instances (i.e., cases) of the problem being addressed, which is the validation set.

As far as error measures are concerned, mean absolute error (MAE) is one of the most frequently used metrics for assessing forecast accuracy and it consists of the mean of the absolute difference, between the total of predicted values ($Pred_i$) for a given output variable and its actual values ($True_i$) for all its *n* observations. Thus, it assesses the deviation in predictive capacity of the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|True_i - Pred_i| \quad (1)$$

Mean absolute percentage error (MAPE) is fundamentally the ratio of the MAE divided by the total of true values. It is the relative variation to those values and it can only be applicable if $True_i > 0$, otherwise the calculation is impossible.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|True_i - Pred_i|}{True_i} \quad (2)$$

Due to the mentioned restriction of MAPE, relative absolute error (RAE) comes into the picture. It is the difference between predicted and true total values as a fraction of the difference between predicted and average total values. This metric enables adjustment to the average values of the variable. The average value (Avg_i) is imputed into the single variations of each sample element which may contribute for increasing accuracy for models with low dispersion. In models with high dispersion, this metric might not be the most suitable because pulling each set of values to its average instead of distributing the total of differences throughout a given *n*, allowing the weights of the differences to be offset within the model, widens enormously the gap in individual sets of values and ends up escalating the total difference. However, it allows assessing predictive capacity when MAPE can't since the average value imputation tackles with the division by zero difficulty. The main advantage of RAE is the ease of interpretations and communication (Armstrong et al, 1992).

$$RAE = \frac{\sum_{i=1}^{n}|True_i - Pred_i|}{\sum_{i=1}^{n}|Avg_i - Pred_i|} \quad (3)$$

Other metrics may be computed to address the issue raised with the RAE. One of such possibilities is to compute a normalized mean absolute error (NMAE), entailing the distribution of the MAE through the difference between the maximum (R_{max}) and minimum (R_{min}) values of the output variable, as shown next:

$$NMAE = \frac{MAE}{R_{max} - R_{min}} \quad (4)$$

MATERIALS AND METHODS

3.1 Approach preamble

When pursuing the employment data mining techniques, one must go through previous and subsequent technical stages in the knowledge discovery process from problem identification and translation into the data mining world to assessing results and possibly repeating the process (Berry et al., 2004). Prediction is a directed data mining task that requires performing all the tasks associated with those different stages, including business understanding, data preparation, modelling, validation and deployment into production or knowledge extraction for decision support (Han et al., 2012).

A comprehensive dataset, including characteristics of sellers and their items, was extracted manually to serve as the base set for the experiments. Sellers represent the problem instances and the set of characteristics comprises both nominal, ordinal and scale features. Moreover, data cleaning, data integration and data transformation, particularly some level of computation to produce new features, were carried out in order to improve the accuracy and efficiency of the mining algorithm (Han et al., 2012; page 83). Issues with missing values were barely registered since problem instances which did not fulfil all the features were immediately eliminated in the data cleaning phase; therefore, there was not a need to implement any method to tackle that problem. Different techniques, tools and metrics are used within the various stages of the process such as using SVM with RBF kernel for modelling, performing a *k*-fold cross-validation, computation of performance metrics MAE, RAE and NMAE, and DSA for assessing feature relevance.

For all the experiments conducted, the R statistical tool (https://cran.r-project.org/) was adopted. R is an open source framework for the development of data analysis solution, with a vast number of enthusiasts and contributors of packages in a wide number of fields of interest (Ihaka & Gentleman, 1996). Moreover, the "rminer" package was adopted as it provides a simple and coherent set of functions for performing data mining tasks such as modelling, model performance evaluation and sensitivity analysis (Cortez, 2010).

3.2 Data preparation

The problem in hands is linked with the shortage of information about which elements among seller and product characteristics have an impact on online sales of smartphones

and how they affect them in a measurable way. The gathered data ensures the reliability of the predictions by way of extracting factual eBay attributes yet its internal validity is tested in following stages of the process using error measures.

Subsequently, one needs to select the appropriate data for the experiments. In this stage, data cleaning was essential in eliminating noise and inconstant data (Han et al, 2012) since all the observations which didn't fulfil all the requirements were immediately removed. The original dataset after data cleaning was composed of 500 observations. Nevertheless, the final dataset included 499 manually extracted reliable observations, which still went through a transformation process prior to mining.

Initially, there were 23 different attributes plus the output variable, which was the number of sales, "prodSales". Those features are listed in Table 1 and identified with source equal to "EBay" in the corresponding column, with Figures 1 to 4 showing the locations on the eBay webpages from where the features were extracted (the respective features' names are identified in the depictions). In the captions of each figure the URL link is also displayed to obtain the webpage identified in each of the figures, for easier reproducibility. Since the R tool was adopted for the experiments, the data types mentioned in Table 1 correspond to R data types (more details on those can be obtained from http://www.dataperspective.info/2016/02/basic-data-types-in-r.html).

Table 1 – List with all the collected and computed features

Feature name	**Source**	**Data type**	**Description**	**Status**
nameSeller	EBay	Character	Name of the seller on eBay	Removed
nrFollowers	EBay	Integer	Total number of followers of the seller	Approved
posR	EBay	Integer	Total number of positive reviews of the seller	Approved
negR	EBay	Integer	Total number of negative reviews of the seller	Approved
neuR	EBay	Integer	Total number of neutral reviews of the seller	Approved
country	EBay	Factor	Country from which the product is sold	Approved
continent	Computed	Factor	Continent of the country	Approved

frItem	EBay	Integer	Feedback rating for the items sold	Approved
frC	EBay	Integer	Feedback rating for the communication	Approved
frST	EBay	Integer	Feedback rating for the shipping time	Approved
frSC	EBay	Integer	Feedback rating for the shipping charges	Approved
diffProd	EBay	Integer	Total products available by the seller	Approved
nrViews	EBay	Integer	Total number of views of the seller	Approved
dateCollection	Computed	Date/time	Date in which data was collected	Removed/Converted
prodType	EBay	Factor	Model of the product	Removed/Converted
segment	Computed	Factor	Assessment based on *prodType*, *brand* and other sources	Approved
brand	EBay	Factor	Brand of the product	Approved
priceMin	EBay	Numeric	Minimum price of the product (€)	Removed/Converted
priceMax	EBay	Numeric	Maximum price of the product (€)	Removed/Converted
nrItems4Sale	EBay	Integer	Number of items in *Buy it now* section	Approved
nrItemsAuction	EBay	Integer	Number of items in *Auction* section	Approved
nrResults4Phone	EBay	Integer	Similar to *diffProd*, except it considers only items under the category of *Cell Phones & Smartphones*	Approved
condition	EBay	Factor	Condition of the product (1=used, 2=refurbished, 3=new)	Approved
moreProdSales	EBay	Integer	Additional sales	Removed
memberSince	EBay	Date/time	Date of membership	Removed/Converted
diffToToday	Computed	Integer	Interval between *dateCollection* and 20th March 2016	Approved
priceAvg	Computed	Numeric	Average of minimum and maximum price (€)	Approved

memberDays	Computed	Integer	Interval between *dateCollection* and *memberSince*	Approved
prodSales	EBay	Integer	Total sales of the product	Approved

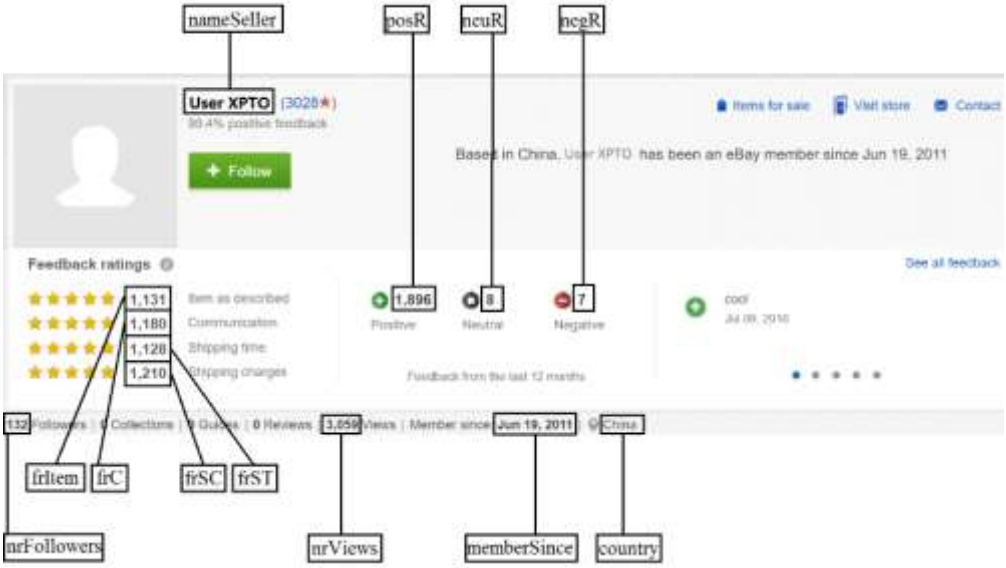


Figure 1 - Locations for the seller's features extracted from eBay (http://www.eBay.com/usr/<user>).

Figure 2 - Locations for the seller's products' features extracted from eBay (http://www.eBay.com/sch/<user>/m.html).

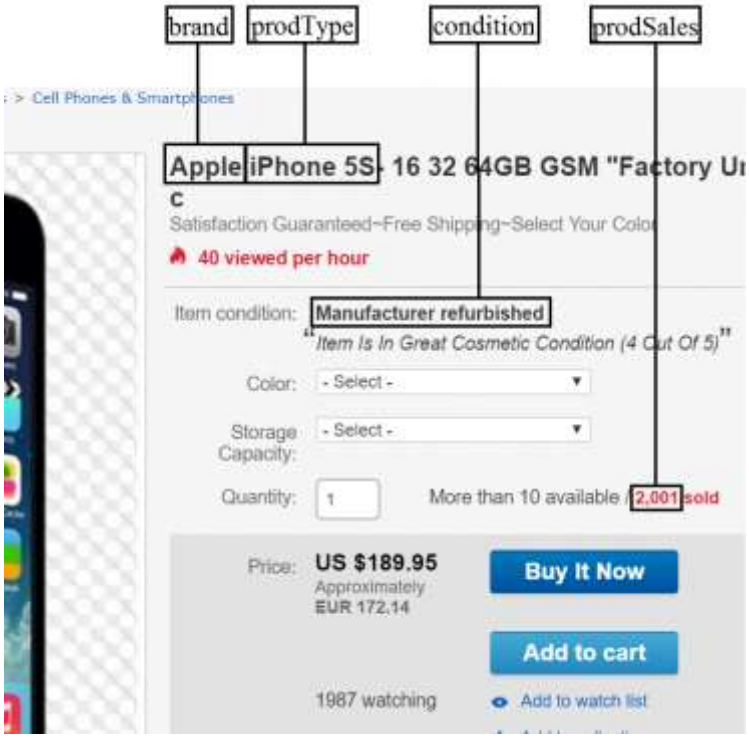


Figure 3 - Locations for the product's features extracted from eBay (http://www.eBay.com/itm/<product>).

~

Figure 4 - Locations for the price's features extracted from eBay (http://www.eBay.com/sch/i.html?_nkw=<product>).

As part of the data preparation process, other additional features were calculated and added to the dataset. These features are also displayed in Table 1 marked with the source "computed". The "dateCollection" registers the date when the data for that record's seller was collected (between February 23 and March 15 2016). The "continent" is another of the features included and was based on the seller's country and using the convention of seven prevailing continents as defined by Lewis (1997): Africa, Antarctica, Asia, Europe, North America, Oceania and South America. Despite the existence of several criteria, the former was chosen because it depicted more precisely the distinctions between Europe and Asia, North America and South America, which can be of value in understanding the effects of the geographical and cultural nature associated with the seller. The "segment" is the result of a categorization based on the smartphone brand and model, the model's release date for covering the issue of older and outdated models over the course of time and reviews from renowned sources such as CNET and GSMArena (Table 2).

Table 2 - Categorization of smartphones' segments.

segment	**brand**	**prodType**	**Release date[3]**	**CNET[4] review**
1	Apple	iPhone 4S	October 2011	8.8
2	Apple	iPhone 5	September 2012	8.7
	ZTE	BoostMax	January 2014	6.7
	BlackBerry	Leap	April 2015	6.6
	Alcatel	OneTouchPopC9	June 2014	N.A.
3	Apple	iPhone 5S	September 2013	8.5
	Huawei	AscendP6	June 2013	N.A.
	Huawei	AscendP7	June 2014	6.7
	ZTE	AXONmini	November 2015	N.A.
	LG	G4	June 2015	8
	Microsoft Mobile	Lumia535	December 2014	6.3
	Microsoft Mobile	Lumia640	March 2015	7.2
	Microsoft Mobile	Lumia640XL	April 2015	N.A.
	Microsoft Mobile	Lumia650	February 2016	6.8
	Huawei	Mate2	January 2014	7.3
	Xiaomi	Mi4c	September 2015	8.7
	Xiaomi	Mi4i	April 2015	N.A.
	HTC	OneMini2	May 2014	7.3
	Alcatel	OneTouchIdol	May 2013	6
	Alcatel	OneTouchIdol3	June 2015	7.6
	Lenovo	VibeShot	June 2015	N.A.

[3] Data retrieved from: http://www.gsmarena.com/
[4] Data retrieved from: http://www.cnet.com/ (N.A. – Not Available)

	Lenovo	VibeZ2Pro	September 2014	N.A.
	Sony Mobile	XperiaC5Ultra	August 2015	N.A.
4	Apple	iPhone 6	September 2014	9
	ZTE	AXONElite	September 2015	N.A.
	HTC	Desire820	November 2014	N.A.
	LG	Gflex2	February 2015	8.3
	Samsung	Note4	October 2014	9
	HTC	OneM8	March 2014	8.7
	Huawei	P8lite	May 2015	6.9
	Samsung	S6	April 2015	8.9
	Sony Mobile	XperiaM5	September 2015	N.A.
	Motorola	XPlay	August 2015	8.4
	Motorola	XStyle	September 2015	7.8
5	Apple	iPhone 6+	September 2014	9
	Apple	iPhone 6S	September 2015	8.9
	Apple	iPhone 6S+	September 2015	9
	Huawei	AscendMate7	October 2014	7.7
	Microsoft Mobile	Lumia950XL	November 2015	7.2
	Huawei	Mate8	November 2015	7.4
	Huawei	MateS	October 2015	7
	Xiaomi	MiNote	January 2015	8
	Huawei	Nexus6P	August 2015	8.4
	HTC	OneA9	November 2015	6.9
	HTC	OneM9	March 2015	8
	Huawei	P8	April 2015	7.9
	BlackBerry	Passport	September 2014	7.3
	Samsung	S6edge	April 2015	9
	Samsung	S6edge+	August 2015	8.8
	Samsung	S7	March 2016	9
	Samsung	S7edge	March 2016	9.1
	LG	V10	October 2015	8.2
	Sony Mobile	XperiaZ5Compact	October 2015	8.7
	Sony Mobile	XperiaZ5Dual	October 2015	7.4
	Sony Mobile	XperiaZ5Premium	November 2015	7.4

Still at the same stage of data preparation, two different features that could not be directly linked to the output variable because they were not quantifiable as an interval were transformed. These are “memberSince” and “dateCollection”, which were converted into “memberDays” – interval between “memberSince” and “dateCollection” – and “diffToToday” – interval between March 20, 2016 (the date when modelling occurred) and “dateCollection” – respectively. Thus, those two features together with

"nameSellers" that was an identification feature, "moreProdSales" that was not pertinent since the registered value was always the same ("N") and "prodType" which contained an unreasonable amount of different categories, a total of 55 within the 499 records, were removed.

Later on the need for a new, more efficient, attribute arose. It was "priceAvg" and it substituted "minPrice" and "maxPrice" through the computation of the average of both. This happened in order to avoid the creation of redundancies because, for most cases, the values registered were the same. Only 46 out of the 499 displayed a difference between both attributes, which was little in most cases. Column "status" from Table 1 reflects the actions taken for each feature, with only the "approved" being included for the modelling stage.

Finally, the 21 different features approved plus the outcome to model ("prodSales") were considered fully functional for proceeding to the next stage, the actual mining of the data.

3.3 Modelling phase

After gathering all the data with adequate methods, this stage is the pinnacle of a data mining project. It is the phase of discovery enabled by the application of suitable intelligent methods, which will subsequently allow extracting knowledge.

Figure 5 shows in a picture the approach followed for this stage. It comprises two main phases. First, the SVM's capabilities of correctly predicting the number of sales for each smartphone's seller are evaluated through a cross-validation scheme with 10-folds. As mentioned before, the RBF kernel is the one in use. For assuring even further the robustness of the model built on the data, the 10-fold cross-validation procedure was run twenty times. In order to evaluate prediction accuracy, three metrics were chosen: MAE, RAE, and NMAE. It should be noted that MAPE was ruled out from this procedure since the dataset contains five records with zero sales, meaning that MAPE cannot be computed for these cases. Furthermore, MAPE distorts the percentage deviation for low values of "prodSales", with this feature ranging from zero to 2,716. Since for each record there are twenty predicted values given the twenty runs of the procedure, the final prediction value for measuring performance is the average of these twenty results.

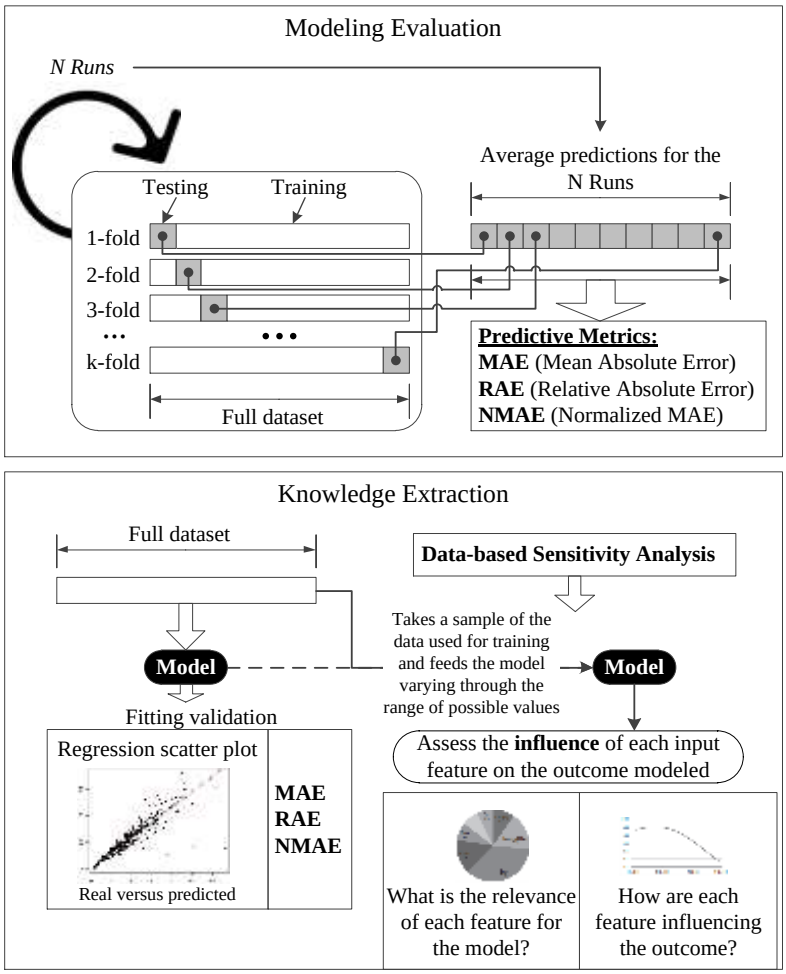


Figure 5 - Scheme with the Modelling Evaluation approach followed.

After assuring that SVM obtains reasonable prediction results, the knowledge extraction phase of the procedure follows. It uses the full dataset to take advantage of the maximum information possible and builds a model based on SVM on top of that data. The validation of fitting the whole dataset is achieved through the three metrics, MAE, RAE and NMAE. Also, for obtaining a visual picture of the deviations of the predictions from the real results, a regression scatterplot is drawn.

Finally, the model built on this second phase is used for knowledge extraction through the DSA. As explained in Section 2.3.2, DSA takes a sample from the dataset used for training the model and then performs an output sensitivity assessment based on varying the input features through their range of possible values. As a result, it becomes possible to assess the influence that each feature has on the number of sales. Two types of valuable knowledge are extracted: the percentage relevance that each feature from the 21 has on

the model and how each of the features affects the number of sales. Such knowledge may provide valuable insights on understanding sellers' performance.

RESULTS AND DISCUSSION

The final step in any data mining project is evaluation and presentation of the findings. An indispensable part of that stage is modelling evaluation that, in case the results are positive, should be followed by knowledge extraction.

4.1 Modelling evaluation

Ascertaining the adequacy of the models involves computing and gauging error measures. Accordingly, MAE was 60.835 whereas RAE was 74.480% and NMAE 2.240%. The substantial discrepancies between the values using different metrics brings out once again the seriousness about choosing the most fitting measure to a given model. NMAE was clearly successful in assessing model adequacy because it is adjusted to the reality of the dataset, i.e., the minimum and maximum values of the output variable. For instance, as n increases it is possible that the difference between R_{max} and R_{min} increases and when using this metric that effect is accounted for without necessarily impairing the results. RAE was not as successful since in models with high dispersion or with a small concentrated cluster of exceptionally high or low values, the average doesn't reflect their broader spectrum. Basically, the more the numerator exceeds the denominator or the closer they are to each other when the numerator is inferior to the denominator, the less the metric will favour model adequacy. Evidently, in models with high dispersion this will lead to adverse results when using RAE. With MAE the results were better than with RAE but worse than with NMAE. This occurs since on one hand the values aren't pulled to the average as with RAE, which is positive in this case, but on the other hand they don't integrate the breadth of values of the output variable as NMAE.

It becomes clear that this particular model is not suitable for higher n values due to the scarcity of comparable observations since there were only five observations with sales above 1000. Nonetheless, the values obtained using the mentioned data mining techniques were good enough to proceed with knowledge extraction.

4.2 Knowledge extraction

In order to show how the employment of different metrics allows fitting validation, the charts below were drawn using MAE (Figure 6) and NMAE (Figure 7) as references for residuals on y-axis at against real sales ($True_i$) on x-axis. The MAE was of 49.981, an

improved result when compared with modelling evaluation. This happens because in the latter the sum of the averages of the subsets is used, i.e. the average of 200 different models, as opposed to the usage of the total dataset in the former, i.e., only 1 model. Thus, RAE also shows improvement at a value of 61.192% along with NMAE at 1.840%.

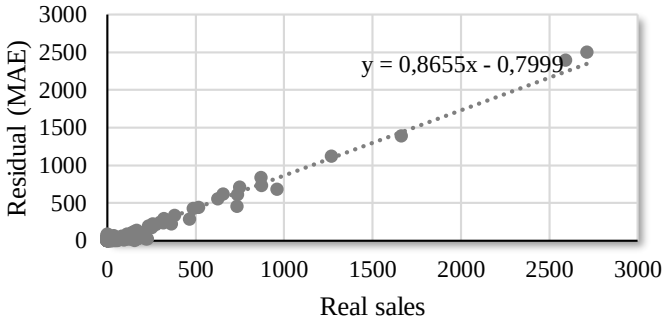


Figure 6 - Regression scatterplot with real sales (x) versus residual with MAE (y).

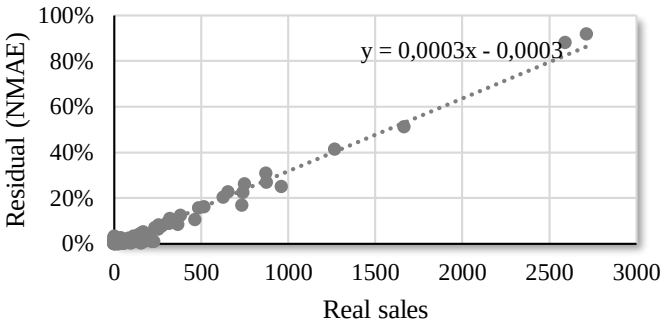


Figure 7 - Regression scatterplot with real sales (x) versus residual with NMAE (y).

Table 3 - Results for the three performance metrics.

	Modelling evaluation	**Knowledge extraction**
MAE	60.84	49.98
RAE	74.48%	61.19%
NMAE	2.24%	1.84%

DSA allowed understanding to which extent the features that fed the SVM algorithm influenced the output variable. Figure 8 exhibits a visual picture of features' relevance while Table 4 shows the percentage values for all the 21 features rounded to the hundredth. Correspondingly, it was discovered that 14 out of the 21 features had a particular influence in the output variable (above 5%), i.e., the number of sales of the smartphone. Their combined contribution to the model is of approximately 91%. The difference between the decomposed contributions of the 14 features was little. The least contributor, the "negR" had an influence of 5% whereas "nrItemsAuction" had an impact of 9%, the highest contribution.

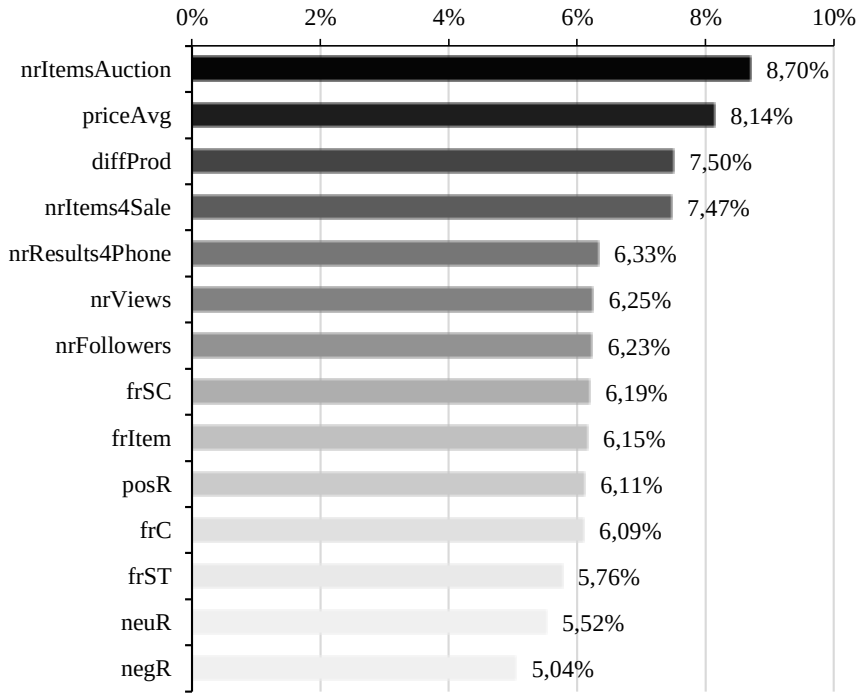


Figure 8 - Features' relevance for modelling sales (shows only values for features with relevance above 5%, rounded to the hundredth).

Table 4 - Features' relevance for modelling sales.

Feature	**Relevance**
nrItemsAuction	8.70%
priceAvg	8.14%
diffProd	7.50%
nrItems4Sale	7.47%
nrResults4Phone	6.33%
nrViews	6.25%

nrFollowers	6.23%
frSC	6.19%
frItem	6.16%
posR	6.11%
frC	6.09%
frST	5.76%
neuR	5.52%
negR	5.04%
diffToToday	2.79%
Segment	1.34%
Country	1.32%
Condition	0.98%
Brand	0.82%
Continent	0.65%
MemberDays	0.64%

The most striking observation that both Figure 8 and Table 4 show is that the five most relevant features, concealing around of 38% of influence, are all related to the assortment of products the seller offers and its management, i.e. showroom-related, including the average price and the range of different products offered by the seller. Organic reach and engagement through "nrViews" and "nrFollowers" respectively also play a role on the number of sales, even though with far less influence than the combined relevance of the five assortment-related features. Such result is aligned with the findings of Moro et al. (2016), which concluded that organic reach and engagement have impact on brand building in social media. Interestingly, the next group of consecutive features in terms of relevance is constituted by seven customer feedback related features, with a combined weight of around 41% of relevance, which are slightly above the product related features in total. This result is a confirmation of previous studies in terms of the influence that customer feedback has on sales (e.g., Kocas & Akkan, 2016). Nevertheless, the top five features remain all showroom-related, relegating individual feedback features for an inferior level.

It was curious that specific product features such as brand and segment along with particular seller features such as country and membership time, barely influenced product sales when compared to the previous ones. The figures mirror the importance of a focused and carefully planned strategy in the background that incentives engagement, promotes reachability and ensures customer satisfaction that is visible through feedback and ratings to the detriment of more specific seller and product features.

The number of items in auction can be traced back to behaviour of bidders striving to get the best deal. As a typical auction website, it is natural that users refine their eBay search looking for items in this category even if they choose another selling format to buy the item. This is reflected in the observed relevance of the input on sales as shown in Figure 9.

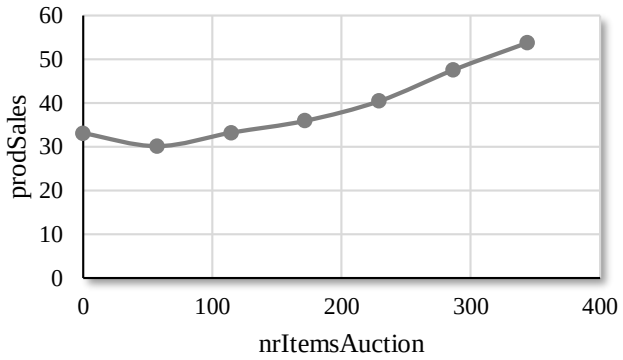


Figure 9 - Impact of number of items in auction on sales.

As expected, "priceAvg" described a fair share of the model as it was the second most significant feature. This happens because price is unquestionably one of the most important marketplace cues (Lichtenstein et al., 1993). Interestingly, it was found that in the particular case of smartphones, product sales plummet until the price level of approximately 1,000 €, after which they start rising (Figure 10). Although it happens many times that product positioning is focused on high-end markets and frequently price is used as a proxy of product quality, as Varian (2014) stated, it also happens often that there is not a direct relation between prices and sales, as sometimes high prices are related to high sales; therefore, the same author argues that continuous experiments with big data and data mining are in demand for obtaining an accurate model.

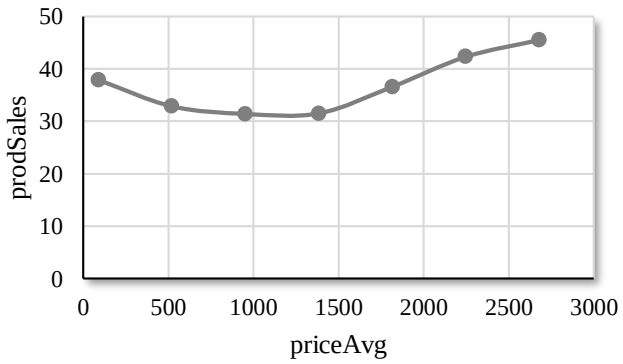


Figure 10 - Impact of average price on sales.

The variety of products sold by a given seller is linked with visibility and assortment selection. The more products are available for sale the more likely it is that views of the seller increase due to inherent exposure. However, if the products fall into many different categories, its assortment might not translate into particular sales of smartphones. Thus, it makes sense that small groups of different products don't have an impact on boosting sales, while after having at least 6206 different products, product sales start increasing (Figure 11).

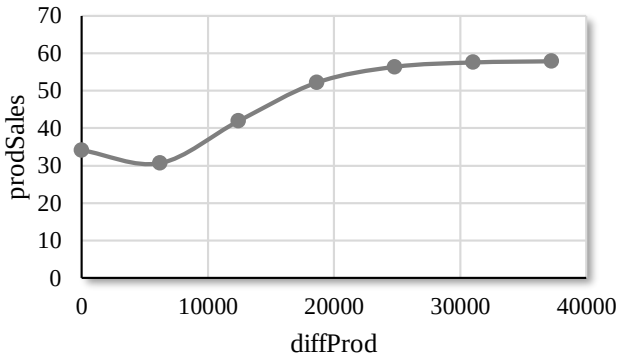


Figure 11 - Impact of assortment on product sales.

The number of items in the "buy it now" listing, "nrItems4Sale", is also linked with visibility since it is tied with search filtering and also assortment management as it reflects a decision taken by the seller regarding the listing of the product. Furthermore, if search is merely product-based, results for items on "buy it now" or in auction can both be presented. Shopping with this type of filtering might, however, be related to preferences for convenience and timeliness since instead of waiting for an auction to end or facing the possibility of losing to another bidder, one can simply buy the product straight away while having access to the same type of information. It is interesting to notice though that after 24,820 items, product sales stabilize at around 60 (Figure 12). However, it should be stressed that only two sellers offer more than that number of products, leading to hypothesize that additional data with sellers offering large number of products would be needed in order to confirm the curve drawn on Figure 12 after the threshold.

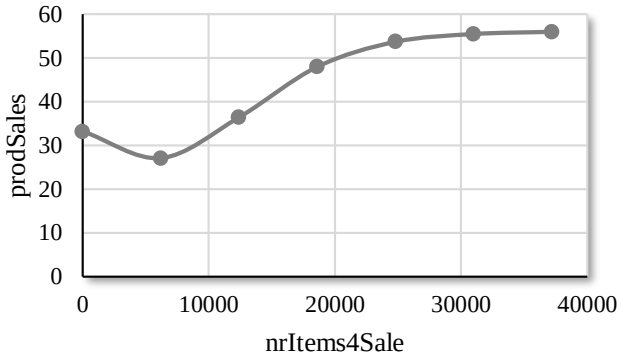


Figure 12 - Impact of number of items in “Buy it now” section on sales.

The influence of the number of results for smartphones shown in Figure 13 reflects the effect of specialization of sellers and therefore their commitment to the category and it is linked with marketing and assortment management strategies. It is natural that people have more trust in specialized sellers of any category than in generalists, which might, in some cases, sell few smartphones in a multitude of products. There are 219 of the sellers within the dataset where the number of smartphones is more than half of the total number of items for sale, while the remaining 280 have a lesser portion of share of smartphones in their stock.

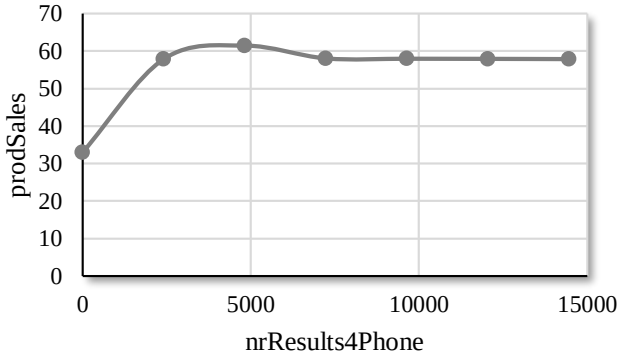


Figure 13 - Impact of specialization on product sales.

The number of views is intrinsically connected with organic reach and it was intriguing that after 186,160 views, product sales continuously dwindled until they reached a plateau (Figure 14). This may be caused by several different factors, for example, as Moro et al. (2016) pointed out for the case of social media, an increase in the number of views may

also provoke some degree of erosion of the seller on eBay. Moreover, there is hardly a direct relation between reachability and market penetration, as other features should be accountable, as observed for the case of "nrResults4Phone".

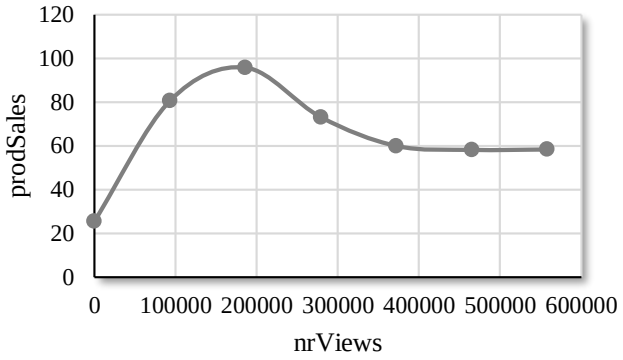


Figure 14 - Impact of number of views on product sales.

Number of followers is linked with reachability and engagement. It is an important source of partial estimates regarding past clients although there might be other reasons for following a seller, which are not covered within the scope of this study. In this case, after 8,294 followers, product sales suffer a slight decay (Figure 15). This shows that after a certain number of followers, there is not any significant increment to sales.

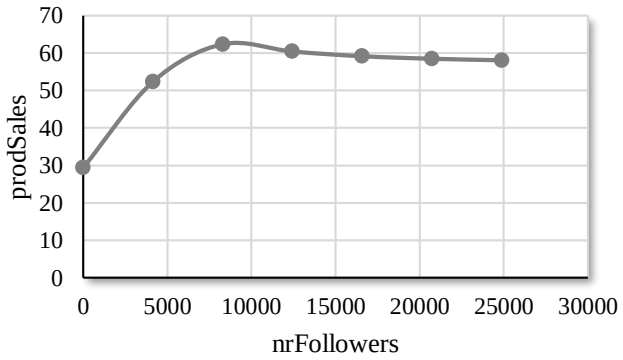


Figure 15 - Impact of number of followers on product sales.

Feedback ratings are often a quite accurate source of varied information about a seller since only after a transaction occurs can the members involved leave a feedback consisting of a short comment and ratings (in the case of the four features considered, only quantitative ratings were included). In this particular case, feedback rating regarding shipping charges was found to be the most significant out of all features within the

typology, even though the difference between the most relevant feature ("frSC") and the least relevant, shipping time ("frST"), is just of 0.43% (Table 4). Such figure may be a result of the worldwide nature of eBay, with registered sellers shipping from around the world, raising the sensitivity that customers have to the values of shipping goods. Figure 16 shows that feedback for communication ("frC") and for the items sold ("frItem") have a similar influence on sales, while both shipping features previously mentioned have also a similar influence between each other. The latter group reveals that shipping feedback results in a more immediate impact on sales, even on a lower number of ratings.

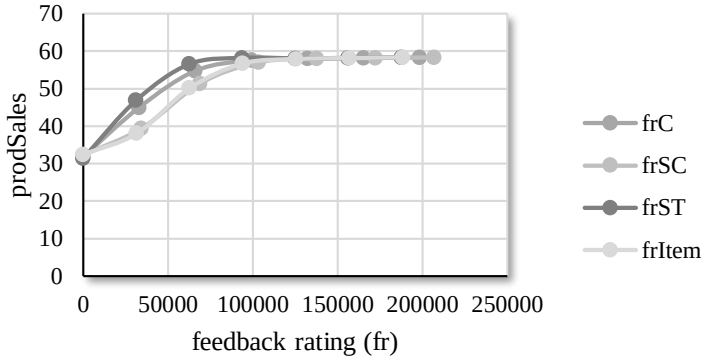


Figure 16 - Impact of feedback rating features on product sales.

The number of positive reviews is another source of valuable insights on successful transactions and it is reasonable to argue that product sales are highly influenced by their value since one can intuitively link a positive review with a positive future response (Vermeulen & Seegers, 2009). It is common that as the number of transactions increases, positive reviews tend to be offset by both neutral and negative reviews, as it was observed throughout the dataset. However, if sellers manage to deliver consistently satisfying products along with the associated service, then product sales are expected to grow (Figure 17).

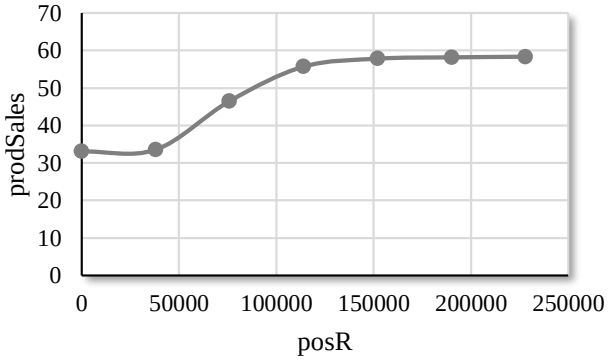


Figure 17 - Impact of the number of positive reviews on product sales.

Within the reviews' typology positive ones have a more gradual impact than neutral reviews and negative, which have higher impact on lower levels of product sales, as it becomes clearly visible in Figure 18. The results reinforce the idea that for smartphones, showroom-related features play a more significant role than customer feedback and, subsequently, reviews that corroborate the initial perception regarding the seller contribute more for sales than neutral or negative ones. On another note, it would be interesting to understand on further research the substantial difference in volume of positive reviews and remaining types which yield more comparable volumes.

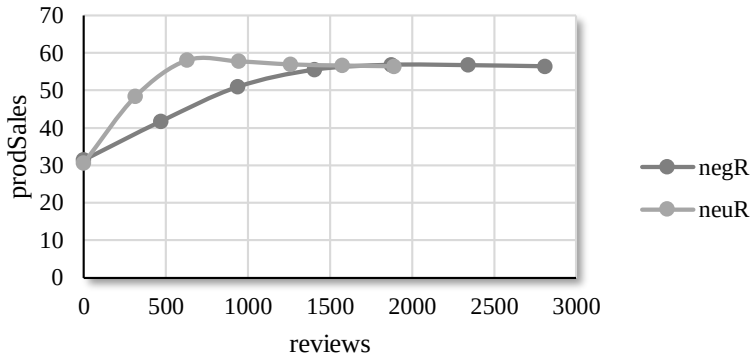


Figure 18 - Impact of the number of negative and neutral reviews on product sales.

CONCLUSIONS

Online marketplaces are currently one of the most thriving forces in retailing, with a huge impact in the global economy. Consequently, eBay, one of the largest online retailers, with a worldwide visibility, is an adequate choice for sellers to promote their products. Furthermore, eBay provides numerous means for customers and users to submit feedback on products and sellers, information that helps to build sellers' reputation. While eBay allows selling any kind of product, the present study is focused on the sales of smartphones, which are sophisticated communication devices with computer capabilities. The smartphones' market is considered one of the most relevant in the information technology field, which has been growing with each new year since the first iPhone was launched by Apple in 2007.

Given the relevance of online sales and, in particular, of eBay and the smartphones' market, the present study is focused on unveiling the features that best characterize the success of smartphones' eBay sellers, measured by the number of sales. In order to succeed in this goal, the approach followed included a data mining project using support vector machines for modelling and data-based sensitivity analysis for extracting knowledge in terms of the relevance of the input features used for modelling the number of sales. The contributions and novelty of the present study lie within two dimensions: on the management perspective, the focus on evaluating the features that best identify a successful seller in terms of the number of sales as opposed to previous studies giving more emphasis on pricing and, on the information science perspective, through the compilation of a previously non-studied dataset including features related with distinct valences such as product (e.g., brand), reachability and engagement (e.g., number of followers), customer feedback (e.g., number of positive reviews), and specific seller information (e.g., the country of origin).

Modelling robustness was tested through a 10-fold cross-validation scheme, executed for twenty times. Model performance was evaluated using three performance metrics: the mean absolute error, the relative absolute error, and the normalized mean absolute error.

The results achieved during the modelling evaluation stage were considered good to proceed with knowledge extraction. Using the data-based sensitivity analysis, it was possible to unveil that the five most relevant features, in a total of around 38%, were all

related to product information. The two features that followed in the relevance rank were related to reachability and engagement, namely the number of views and the number of followers. Next, seven customer feedback related features appeared, concealing a total of around 41% of relevance, in the eighteenth to the fourteenth position. The discovery that the individual features related to customer feedback are less relevant than those related to the showroom and reachability/engagement is interesting. Nevertheless, it should be noted that the total seven features for feedback and reviews also play a role as a whole since they conceal the highest percentage of relevance (41%).

In essence, sales of smartphones on eBay are mostly influenced by showroom-related features, which reflect the underlying marketing and assortment management strategies ("nrItemsAuction"; "priceAvg"; "diffProd"; "nrItems4Sale"; "nrResults4Phone") and are deeply enhanced by reach ("nrViews"), engagement ("nrFollowers") while being supported by access to several sources of feedback and reviews about the seller.

By further taking advantage of the sensitivity analysis, it was possible to observe how each of the most relevant features affected the number of sales. For example, the most relevant feature, i.e., the number of items the seller has in auction influences the number of sales in a linear proportion, i.e., the more items in auction, the higher the number of sales. Such result may derive from the fact that a seller with a lot of items in auction benefits from customers who do not want to wait for the outcome of an auction and instead choose to buy the item directly from the seller.

The present study has some limitations that may be addressed in future research. First, only eBay data was used for the experiments. While eBay is one of the largest online retailers, other huge players have risen in the past recent years; most notably, Taobao, from China. Therefore, in the future, a much larger dataset could be compiled from different sources, namely through the usage of web scrapping tools that can automatically extract the features from the different webpages. Additionally, other features could be devised and tested, in order to enrich the model's knowledge about the number of sales.

REFERENCES

Aldrich, M. 2011. Online Shopping in the 1980s. ***IEEE Annals of the History of Computing***, 33(4), 57-61.

Alexa. ***The top 500 sites on the web***. Retrieved from here (March 2016)

Armstrong, J. S., & Collopy, F. 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. ***International journal of forecasting***, 8(1), 69-80.

Bajari, P., & Hortacsu, A. 2004. Economic insights from internet auctions. ***Journal of Economic Literature***, *42*(2), 457-486.

Berry, T. A., McKeen, T. R., Pugsley, T. S., & Dalai, A. K. 2004. Two-dimensional reaction engineering model of the riser section of a fluid catalytic cracking unit. ***Industrial & engineering chemistry research***, 43(18), 5571-5581.

Bilgihan, A., Kandampully, J. & Zhang, T. 2016. Towards a unified customer experience in online shopping environments: Antecedents and outcomes. ***International Journal of Quality and Service Sciences***, 8(1), 102-119.

Black, G. S. 2007. Consumer demographics and geographics: Determinants of retail success for online auctions. ***Journal of Targeting, Measurement and Analysis for Marketing***, 15(2), 93-102.

Bruce, N., Haruvy, E., & Rao, R. 2004. Seller rating, price, and default in online auctions. ***Journal of Interactive Marketing***, 18(4), 37-50.

Brynjolfsson, E., Hu, Y. J., & Smith, M. D. 2006. From niches to riches: Anatomy of the long tail. ***Sloan Management Review***, 47(4), 67-71.

Cao, P., Fan, M., & Liu, K. 2015. Optimal dynamic pricing problem considering patient and impatient customers' purchasing behaviour. ***International Journal of Production Research***, 53(22), 6719-6735.

Cheema, A., Leszczyc, P. T. P., Bagchi, R., Bagozzi, R. P., Cox, J. C., Dholakia, U. M. & Sunder, S. 2005. Economics, psychology, and social dynamics of consumer bidding in auctions. ***Marketing Letters***, 16(3-4), 401-413.

Chen, H., Chiang, R. H. & Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. ***MIS Quarterly***, 36(4), 1165-1188.

Chen, J., Chen, J. E., Goh, K. Y., Xu, Y. C., & Tan, B. C. 2014. When do sellers bifurcate from Electronic Multisided Platforms? The effects of customer demand, competitive intensity, and service differentiation. ***Information & Management***, 51(8), 972-983.

Chen, J., Chen, X., & Song, X. 2002. Bidder's strategy under group-buying auction on the Internet. ***Systems, Man and Cybernetics, Part A: Systems and Humans***, IEEE Transactions on, 32(6), 680-690.

Chen, P. Y., Wu, S. Y., & Yoon, J. 2004. The impact of online recommendations and consumer feedback on sales. ***ICIS 2004 Proceedings***, 58.

Cheung, C. M., Chan, G. W., & Limayem, M. 2005. A critical review of online consumer behavior: Empirical research. ***Journal of Electronic Commerce in Organizations***, 3(4), 1-19.

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. 2015. Predicting consumer product

demands via Big Data: the roles of online promotional marketing and online reviews. ***International Journal of Production Research***, 1-15.

Cortes, C., & Vapnik, V. 1995. Support-vector networks. ***Machine learning***, 20(3), 273-297.

Cortez, P. 2010. ***Data mining with neural networks and support vector machines using the R/rminer tool***. In Industrial Conference on Data Mining (pp. 572-583). Springer Berlin Heidelberg.

Cortez, P., & Embrechts, M. J. 2011. Opening black box data mining models using sensitivity analysis. In ***Computational Intelligence and Data Mining (CIDM)***, 2011 IEEE Symposium on (pp. 341-348). IEEE.

Cortez, P., & Embrechts, M. J. 2013. Using sensitivity analysis and visualization techniques to open black box data mining models. ***Information Sciences***, 225, 1-17.

Diebold, F. X., & Mariano, R. S. 2012. Comparing predictive accuracy. ***Journal of Business & Economic Statistics***, 13(3), 253-263.

EBay. ***Bidding overview***. Retrieved from here (28th April 2016)

Einav, L., Levin, J., Popov, I., & Sundaresan, N. 2014. Growth, adoption, and use of mobile E-commerce. ***The American Economic Review***, 104(5), 489-494.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. 1996. From data mining to knowledge discovery in databases. ***AI magazine***, 17(3), 37.

Fisher, M. L. 1997. What is the right supply chain for your product? ***Harvard Business Rreview***.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. ***Annals of eugenics***, 7(2), 179-188.

Friedman, J., Hastie, T., & Tibshirani, R. 2001. ***The elements of statistical learning (Vol. 1)***. Springer, Berlin: Springer series in statistics.

Goes, P., Tu, Y., & Tung, Y. A. 2013. Seller heterogeneity in electronic marketplaces: A study of new and experienced sellers in eBay. ***Decision Support Systems***, 56, 247-258.

Gregg, D. G., & Scott, J. E. 2008. A typology of complaints about eBay sellers. ***Communications of the ACM***, 51(4), 69-74.

Grewal, D., Janakiraman, R., Kalyanam, K., Kannan, P. K., Ratchford, B., Song, R., & Tolerico, S. 2010. Strategic online and offline retail pricing: a review and research agenda. ***Journal of Interactive Marketing***, 24(2), 138-154.

Gunn, S. R. 1998. Support vector machines for classification and regression. ***ISIS technical report***, 14.

Han J., Kamber A. & Pei J.2012. ***Data mining: Concepts and Techniques***. 3rd Edition, Elsevier, USA.

Hanna, M. 2004. ***Data mining in the e-learning domain***. Campus-wide information systems, 21(1), 29-34.

Haucap, J., & Heimeshoff, U. 2014. Google, Facebook, Amazon, eBay: Is the Internet driving competition or market monopolization? International Economics and Economic Policy, 11(1-2), 49-61.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. 1998. Support vector machines. ***Intelligent Systems and their Applications***, IEEE, 13(4), 18-28.

Hemp, P. 2006. Are you ready for e-tailing 2.0? ***Harvard Business Review***. June 2006, 28.

Hess, C. M., & Kemerer, C. F. 1994. Computerized loan origination systems: an industry case study of the electronic markets hypothesis. ***MIS Quarterly***, 251-275.

Huang, X., & Finch, B. J. 2010. Satisfaction and dissatisfaction in online auctions: an empirical analysis. ***International Journal of Quality & Reliability Management***, 27(8), 878-892.

Hyndman, R. J., & Koehler, A. B. 2006. Another look at measures of forecast accuracy. ***International journal of forecasting***, *22*(4), 679-688.

IBM. 2016. Guide to Retail Technology Trends.

Ihaka, R., & Gentleman, R. 1996. R: a language for data analysis and graphics. ***Journal of Computational and Graphical Statistics***, 5(3), 299-314.

Jayaraman, V., & Baker, T. 2003. The Internet as an enabler for dynamic pricing of goods. ***Engineering Management***, *IEEE Transactions on*, *50*(4), 470-477.

Kellerman, A. 2010. Mobile broadband services and the availability of instant access to cyberspace. ***Environment and Planning A***, 42(12), 2990-3005.

Khodakarami, F., & Chan, Y. E. 2014. Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. ***Information & Management***, 51(1), 27-42.

Kocas, C., & Akkan, C. 2016. How Trending Status and Online Ratings Affect Prices of Homogeneous Products. ***International Journal of Electronic Commerce***, 20(3), 384-407.

Kotler P. and Keller K. 2012. ***Marketing Management***. 14th Edition, Prentice Hall, USA.

Kusiak, A., & Smith, M. 2007. Data mining in design of products and production systems. ***Annual Reviews in Control***, 31(1), 147-156.

Laudon, K. C., & Traver, C. G. 2007. ***E-commerce***. 10th Edition, Pearson/Addison Wesley

Lee, C. S., Ho, J. C., & Hsu, C. F. 2015. Creating value in global innovation networks: A study of smartphone industry. In ***Management of Engineering and Technology*** (PICMET), 2015 Portland International Conference on (pp. 755-760). IEEE.

Lee, S., & Choeh, J. Y. 2014. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. ***Expert Systems with Applications***, 41(6), 3041-3046.

Leeflang, P. S., Verhoef, P. C., Dahlström, P., & Freundt, T. 2014. Challenges and solutions for marketing in a digital era. ***European management journal***, 32(1), 1-12.

Lewis, M. W. 1997. The myth of continents: A critique of metageography. ***University of California Press***.

Lichtenstein, D. R., Ridgway, N. M., & Netemeyer, R. G. 1993. Price perceptions and consumer shopping behavior: a field study. ***Journal of Marketing Research***, 30(2), 234-245.

Liu, S., & Lu, C. 2015. Cultural tourism O2O business model innovation-case analysis based on CTRIP. ***In Logistics, Informatics and Service Sciences (LISS)***, 2015 International Conference on (pp. 1-6). IEEE.

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., & Kallos, G. 2008. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. ***Journal of Wind Engineering and Industrial Aerodynamics***, 96(12), 2348-2362.

Marlin, B. 2004. Modelling user rating profiles for collaborative filtering. ***Advances in Neural Information Processing Systems***, 16, 627–634.

MathWorks. 2016. ***Machine Learning with MATLAB***. Retrieved from here (14th April 2016)

Moro, S., Cortez, P., & Rita, P. 2014. A data-driven approach to predict the success of bank telemarketing. ***Decision Support Systems***, 62, 22-31.

Moro, S., Cortez, P., & Rita, P. 2015. Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns. ***Neural Computing and Applications***, 26(1), 131-139.

Moro, S., Rita, P., & Vala, B. 2016. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. ***Journal of Business Research***, 69(9), 3341-3351.

Pal, K., & Saini, J. 2014. A Study of Current State of Work and Challenges in Mining Big Data. ***International Journal of Advanced Networking Applications***, Special Issue, 73-76.

Park, J., Lennon, S. J., & Stoel, L. 2005. On-line product presentation: Effects on mood, perceived risk, and purchase intention. ***Psychology & Marketing***, 22(9), 695-719.

Pearce, K. E. and Rice, R. E. 2013. Digital divides from access to activities: Comparing mobile and personal computer Internet users. ***Journal of Communication***, 63(4), 721-744.

Pinker, E. J., Seidmann, A., & Vakrat, Y. 2003. Managing online auctions: Current business and research issues. Management science, 49(11), 1457-1484.

Poelker. 2013. ***Smartphones, big data, storage and you***. Retrieved from ComputerWorld, here (26th June 2016).

Ramdas, K. 2003. Managing product variety: An integrative review and research directions. ***Production and operations management***, 12(1), 79-101.

Ratchford, B. T. 2009. Online pricing: review and directions for research. ***Journal of Interactive Marketing***, 23(1), 82-90.

Refaeilzadeh, P., Tang, L., & Liu, H. 2009. Cross-validation. ***In Encyclopedia of database systems*** (pp. 532-538). Springer US.

Saltelli, A., Chan, K., & Scott, E. M. (Eds.). 2000. Sensitivity analysis (Vol. 1). ***New York: Wiley***.

Schölkopf, S. P., Vapnik, V. & Smola, A. J. 1997. Improving the accuracy and speed of support vector machines. ***Advances in neural information processing systems***, 9, 375-381.

Sharda, R., Delen, D. & Turban, E. 2014. ***Business Intelligence and Analytics: Systems for Decision Support***. Pearson Education.

Smola, A. J., & Schölkopf, B. 1998. On a kernel-based method for pattern recognition, regression, approximation, and operator inversion***. Algorithmica***, 22(1-2), 211-231.

Statista. Global market share held by leading smartphone vendors from 4th quarter 2009 to 2nd quarter 2016. Retrieved from here (September 2016).

Statista. Number of internet users worldwide. Retrieved from here (24th March 2016).

Statista. Number of smartphones sold to end users worldwide from 2007 to 2015 (in million units. Retrieved from here (24th March 2016)

Statista. Share of internet users who have ever purchased products online as of October 2015, by category. Retrieved from here (24th March 2016).

Statista. Share of smartphone shipments worldwide from 2012 to 2015, by vendor. Retrieved from here (24th March 2016)

Statista. Statistics and facts about Smartphones. Retrieved from here. (24th March 2016).

Tadelis, S. 2016. 19 two-sided e-commerce marketplaces and the future of retailing. Handbook on the Economics of Retailing and Distribution. ***In Basker, E. (Ed.), Handbook on the Economics of Retailing and Distribution***. Edward Elgar Publishing.

The Guardian. ***Online shopping on mobiles overtakes desktop for first time***. Retrieved from here (17th March 2016).

Varian, H. R. 2014. Beyond big data. Business Economics, 49(1), 27-31.

Venkatesan, R., Mehta, K., & Bapna, R. 2006. Understanding the confluence of retailer characteristics, market characteristics and online pricing strategies. ***Decision Support Systems***, 42(3), 1759-1775.

Vermeulen, I. E., & Seegers, D. 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. ***Tourism management***, 30(1), 123-127.

Wilcox, R. T. 2000. Experts and amateurs: The role of experience in Internet auctions. ***Marketing Letters***, 11(4), 363-374.

Xu, M., & Ye, Q. 2015. Reputation and pricing strategies in online market. ***Proceedings on the Wuhan International Conference on e-Business (WHICEB)***, 678-684.

Yan, Z., Yan, L., & Leboulanger, M. 2009. National and Cultural Differences in the C2C Electronic Marketplace: An Investigation into Transactional Behaviors of Chinese, American, and French Consumers on eBay. ***Tsinghua Science and Technology***, 14(3), 383-389.

Ye, Q., Xu, M., Kiang, M., Wu, W., & Sun, F. 2013. In-depth analysis of the seller reputation and price premium relationship: A comparison between eBay us and Taobao china. ***Journal of Electronic Commerce Research***, 14(1), 1.

Yen, C. H., & Lu, H. P. 2008. Factors influencing online auction repurchase intention. ***Internet Research***, 18(1), 7-25.

Yu, X., Liu, Y., Huang, X., & An, A. 2012. Mining online reviews for predicting sales performance: A case study in the movie domain. ***IEEE Transactions on Knowledge and Data engineering***, 24(4), 720-734.

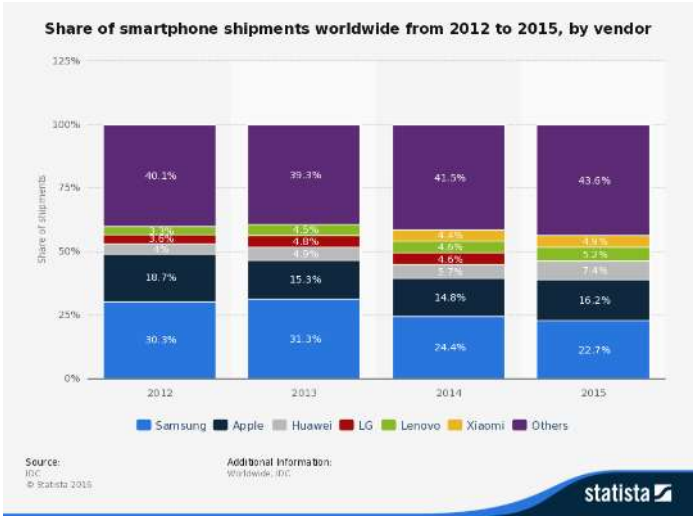
Zhang, J., Farris, P. W., Irvin, J. W., Kushwaha, T., Steenburgh, T. J., & Weitz, B. A.

2010. Crafting integrated multichannel retailing strategies. ***Journal of Interactive Marketing***, 24(2), 168-180.

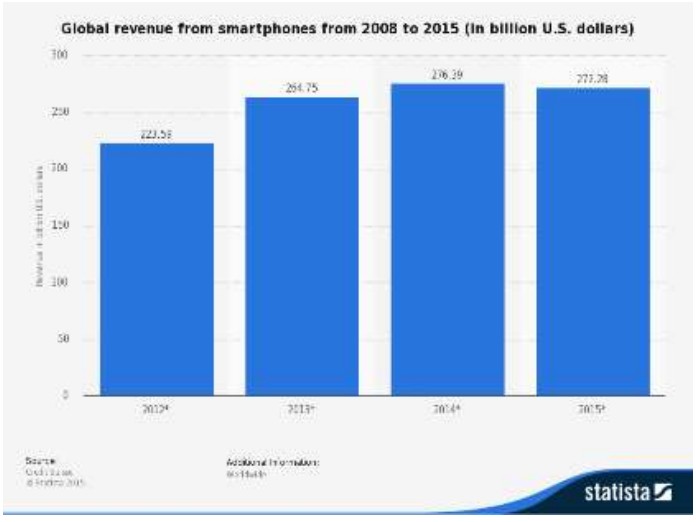
Zhu, Y., Li, Y., & Leboulanger, M. 2009. National and cultural differences in the C2C electronic marketplace: An investigation into transactional behaviors of Chinese, American, and French consumers on eBay. ***Tsinghua Science & Technology***, 14(3), 383-389.

ANNEXES

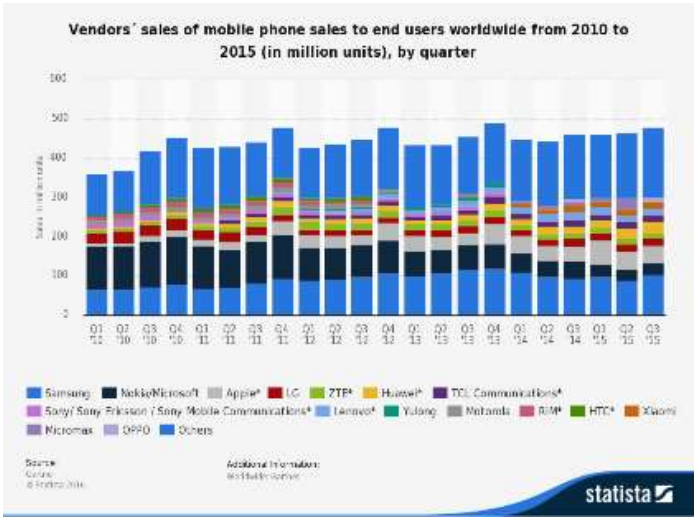
Annex 1 - Share of smartphone shipments worldwide from 2012 to 2015, by vendor (Statista, 2016).



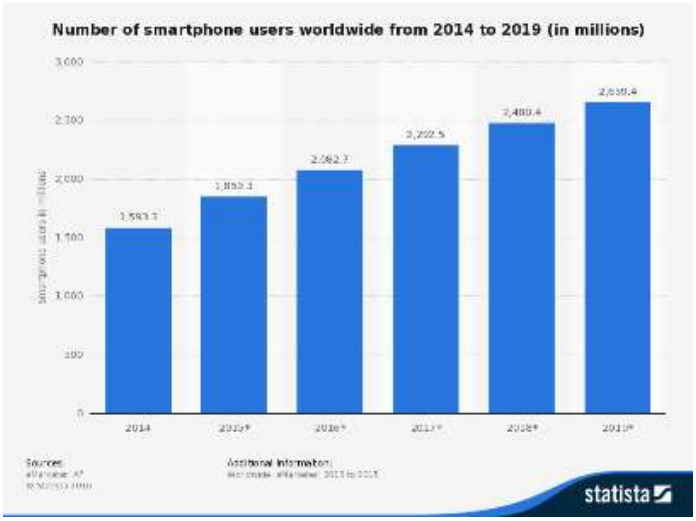
Annex 2 - Global revenue from smartphones from 2008 to 2015 (in billion U.S. dollars) (Statista, 2016).



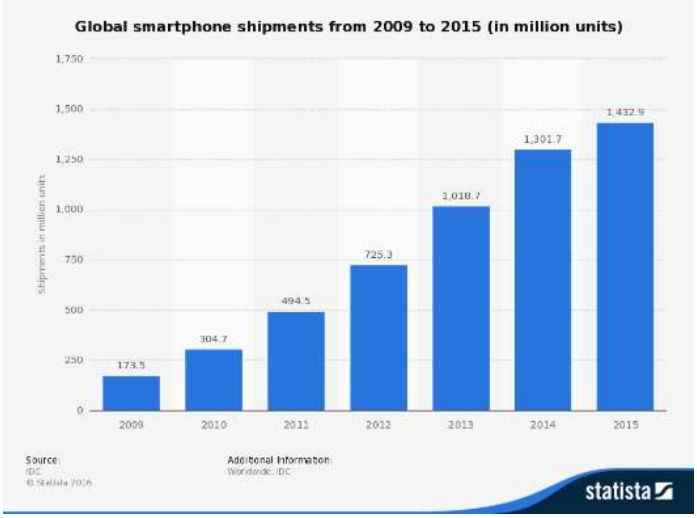
Annex 3 - Vendors' sales of mobile phones to end users worldwide from 2010 to 2015 (in million units), by quarter (Statista, 2016).



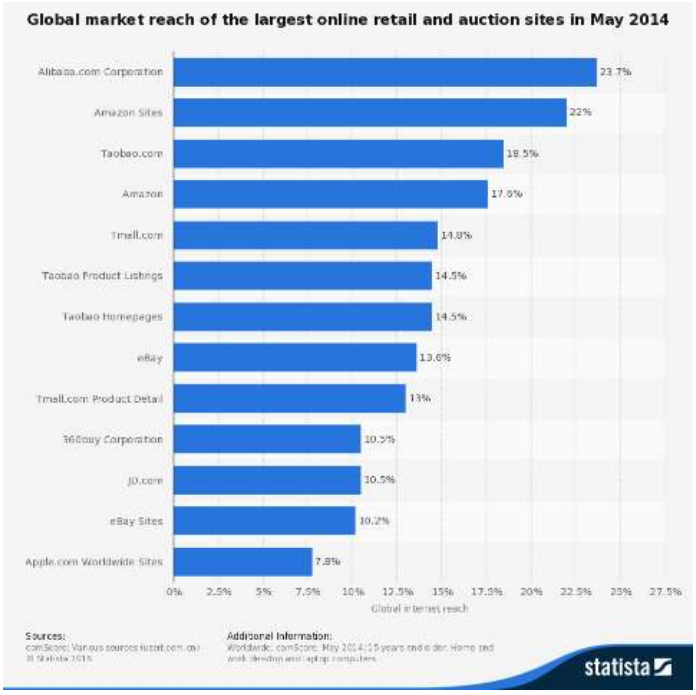
Annex 4 - Number of smartphone users worldwide from 2014 to 2019 (in millions) (Statista, 2016).



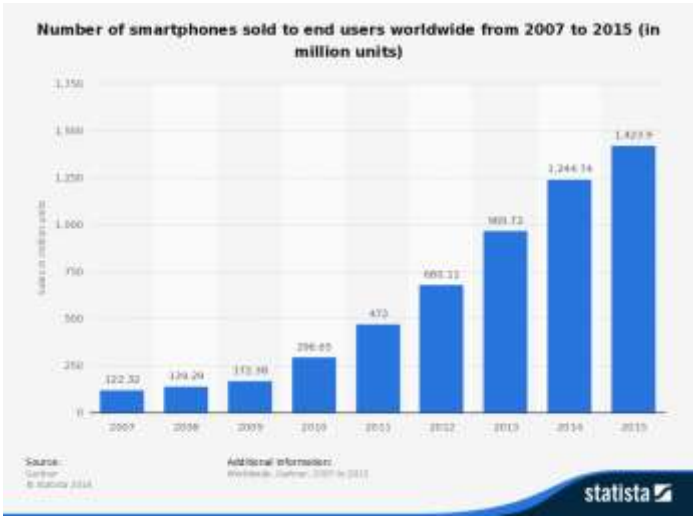
Annex 5 - Global smartphone shipments from 2009 to 2015 (in millions units) (Statista, 2016).



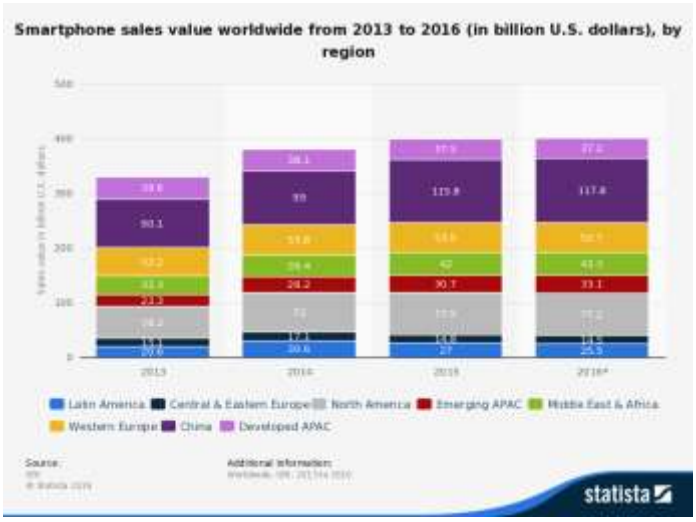
Annex 6 - Global market reach of the largest online retail and auction sites in May 2014 (Statista, 2016).



Annex 7 - Number of smartphones sold to end users worldwide from 2007 to 2015 (in million units) (Statista, 2016).



Annex 8 - Smartphone sales value worldwide from 2013 to 2016 (in billion U.S. dollars), by region (Statista, 2016).



Annex 9 - Global average selling price of smartphones from 2010 to 2019 (in U.S. dollars) (Statista, 2016).