Business Intelligence Implementation on the Marketing of Real Estate

Douglas S. Acevedo Martínez
Master of Engineering in Computer Engineering
Advisor: Dr. Nelliud Torres
Electrical and Computer Engineering and Computer Science Department
Polytechnic University of Puerto Rico

Abstract — *This research identify the Business Intelligence's components and techniques that could be implemented on different industries. Also, I review literature available in relation of the implementation of Business Intelligence systems within the Real Estate Industry. This study conducted a data mining analysis using an available public data from 2007 until 2017. Using this data mining, I seek to follow the Business Intelligence principle of translating data into knowledge. The analysis identify trends that confirm the valuable information gain by these techniques and these were translated into strategies that could be implemented for improving the quality of services provided by real estate professionals. To conclude the paper, I used data warehouse containing the loan application data to developed different models to predict the loan amount applications and the prediction model seek to serve as an example that could be used to translate the results on potential strategies for the real estate professional.*

Key Terms — *The following terms are relevant terms used on this article: Business Intelligence, Data Mining, Real Estate, Predictive Model.*

INTRODUCTION

During the early 1990s, an analyst working for Gartner Group, named Howard Dressner, introduced the concept of Business Intelligence (BI). As mentioned by Watson and Wixon (2007), BI is currently the top-most priority of many Chief Information Officers and has become a strategic initiative for driving business effectiveness and innovation [1]. Ranjan (2005) stated that many companies are adopting BI tools and systems to learn from the past and forecast the future [2].

Based on Trieu (2016), much of the research on BI has examined the ability of BI systems to help organizations address challenges and opportunities [3]. BI enables companies and organization to make well informed business decisions and acquire competitive advantages [2]. BI systems are now used extensively in many areas of business that involve making decisions to create value [3]. Most organizations have been impacted by the BI revolution and their strategies have changed business processes to make them more efficient and to create a new competitive advantage for these new business practices (Poulis and Thamir, 2015) [4]. At the same time as most organizations start their BI, many struggle to align their technology approach to BI and business objectives, and deliver solutions that fail to meet business needs.

Businesses are leveraging their data asset aggressively by deploying and experimenting with more sophisticated data analysis techniques to drive business decisions and deliver new functionality such as personalized offers and services to customers [5]. Although BI has been around for a long time, there are significant opportunities created by the availability of big data and advances in machine learning [6]. Today, it is difficult to find a successful enterprise that has not leveraged BI technology for their business [5].

The objective of the Real Estate Agent is to facilitate the real estate transaction by marketing the seller's property, search for a property that meets the requirements of the buyer and providing consultation to buyer and/or seller during every step of the process. In the era of information, the Real Estate Agents has not capitalized on the opportunity of using data, applications and technology to increase services value and performance.

At the forefront of technology, tech and software companies are battling to create Artificial Intelligence that will begin to not only automate

parts of sales, but also allow businesses to make better decisions than people, and real estate is just one of the industries poised for disruption [7].

LITERATURE REVIEW

Business Intelligence (BI) is a term used to describe the use of all the comprehensive data owned by a business to gain a depth knowledge to make effective and quality decisions. Ranjan (2005) defines BI as the application of a set of methodologies and technologies to improve enterprise operation effectiveness, support management/decision to achieve competitive advantages [2]. According to Adelman, Moss and Barbusinski (2002), BI is a term that encompasses a broad range of analytical software and solutions for gathering, consolidating, analyzing and providing access to information to let users make better business decisions [8]. Trieu define the term BI as the set of concepts and methods based on fact-based support systems for improving decision making [3]. While BI system are defined as a system compromised of both technical and organizational elements that presents historical information to its users for analysis, query and reporting, to enable effective decision-making and management support, to increase the performance of business processes [3].

The traditional BI solutions rely in ETL (Extraction, Transformation and Load) data processes to form a data warehouse, where the BI tools transform information to knowledge. Based on Ranjan (2005), the primary activities include gathering, preparing and analyzing data to achieve the ultimate objective of BI of improving the timeliness and quality of information [2]. Therefore, BI is a combination of tools to transform historical data into knowledge and methodologies that could be used to conduct the appropriate analysis for the decision process, to establish business strategies and/or get competitive advantage.

Business intelligence converts data into useful information and into knowledge [9]. BI acquire data from various different systems within the organization (Fig 1).
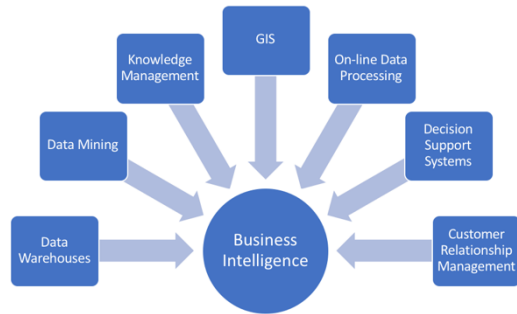

Figure 1
Examples of BI Data Inputs

Therefore, the components of Business Intelligence are [2] [9]:

- On-line analytical processing (OLAP),
- Real-time data metrics and data,
- Data warehouse and data marts,
- Data mining,
- Data sources,
- Data visualization,
- Automatic learning and refinement,
- Advanced analytics,
- Proactive alerting with automatic recipient determination, and
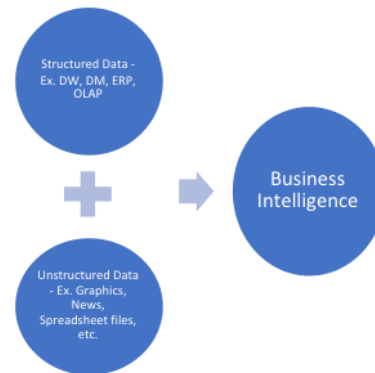- Seamless follow-through workflow.

Figure 2
Types of BI Data

Considering that the BI obtained data from a diverse amount of systems, it should be required to deal with every type of data. As illustrated on Figure 2, the different types of data can be classified as structured data and semi-structured

data. Structured data is data that contain relations or is organized on a database, while unstructured data do not fit on this description [9].

In the past, business analysis used structured data, and reports were generated by answering predetermined questions. In contrast, nowadays this is a more interactive process using relational data and tools that facilitate the data analysis. Various research papers analyzed or proposed business intelligence architectures for structured data, but most of them consider the same elements:

- External data sources or operational data store on databases
- Systems to Extract, Transform and Load (ETL) the data
- Data Warehouses and Data Mart
- Tools and Analytics
- User/Client Interface of Reporting

The process of business intelligence, that involves collecting, extracting, transforming, and loading data (ETL) for data mining and analysis, is strongly influenced by the practices related to strategy development [10].

Data Mining

Data mining techniques were developed by 1980s and it's considered an area of Business Intelligence. Data mining enables in-depth analysis of data including the ability to build predictive models [5]. A series of data-mining techniques are illustrated in Table 1 [2].

The approach of data mining is to select a subset of data from the data warehouse, perform sophisticated data analysis on the selected subset of data to identify key statistical characteristics, and to then build predictive models that are deployed in the operational database [5].

According Branscombe (2016), a lot of business want to get the pulse of their business using BI but dashboards show things that have already happened, so BI tools goes beyond to more sophisticated methods like statistics, descriptive and predictive data mining, machine learning,

simulation and optimization that looks for trends and patterns in the data [7].

Table 1
Current BI Techniques

TECHNIQUE	DESCRIPTION
Predictive modeling	Predict value for a specific data item attribute
Characterization and descriptive data mining	Data distribution, dispersion and exception
Association, correlation, causality analysis	Identify relationships between attributes
Classification	Determine to which class a data item belongs
Clustering and outlier analysis	Partition a set into classes, whereby items with similar characteristics are grouped together
Temporal and sequential pattern analysis	Trend and deviation, sequential patterns, periodicity
OLAP (On Line Analytical Processing)	OLAP tools enable users to analyze different dimensions of multidimensional data
Model Visualization	Making discovered knowledge easily understood using charts, plot, histograms, and other visual means
Exploratory Data Analysis (EDA)	Explores a data set without a strong dependence on assumptions or models; goal is to identify patterns in an exploratory manner

Business Intelligence in the Real Estate Industry

Most of the industries are always looking for ways to increase efficiency and use technologies to help them save time and money on their businesses. The real estate industry is not the exception, and professionals are increasingly looking for ways to reduce manual processes and use technologies such as machine learning to:

- Identify key terms and red flags,
- Helps to reduce manual searches, and

- Automate physical processes.

New research has revealed that machine learning could help reduce the need for paper documents and make real estate data more accessible [11].

According to Berger (2017), there are a number of ways that BI and technologies could support the real estate industry. Some examples are the following [12]:

- Streamline processes for tenants, landlords and property managers.
- Help property managers find new tenants, locate cost-efficient vendors, and proactively recommend maintenance and management tasks.
- Tenants can pay rent online and report issues through a convenient chatbot feature.
- Bots can answer questions on square footage, leasing terms, and other topics of interest. If it doesn't have an answer, it will transfer the user to a human agent.
- Match the buyers with spaces after learning their preferences.
- Uses a variety of data sources and models to create an approximate value for residential properties [13].

Experienced real estate agents, after diving into their customers' journey and data to make descriptive assumptions about their clientele and the market, try to predict [7]:

- average time to close on a home
- seasonal or market cycles
- best time to call or send an email
- where to target particular marketing media for optimal performance
- what tone or language to use with a particular customer.

Based on the study conducted by Hoffmeister (2017), the real estate professional's overwhelming requirement is for technology that can assist real estate professionals in their ability to swiftly assimilate knowledge and insights from a huge range of data and documents [11].

IBM first structured the data analysis landscape, from first descriptive to predictive to finally prescriptive analytics [7], as illustrated in figure 3.
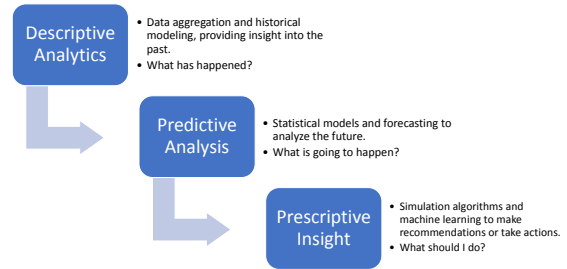

Figure 3
IBM's Data Analysis Landscape [14]

Although the rise of machine learning might be inevitable, agents who don't learn how to leverage automation and AI to grow their business will find themselves out of a job, while salespeople who do adopt technology are poised to inherit the future of real estate [7]. It's important to understand that three conditions must be met before applying machine learning to a problem though [14]:

- There must be a pattern in the input data to arrive at a conclusion.
- There has to be enough data to apply machine learning to a problem.
- Humans are unable to formulate a mathematical expression that describes the behavior of the problem.

According to Levy-Yurista, by 2020, the number of users of modern business intelligence and analytics platforms that are differentiated by augmented data discovery capabilities will grow at twice the rate and deliver twice the business value of those that are not [15].

PROBLEM STATEMENT

For over the past 15 years, the real estate marketing strategies used by Puerto Rico's market have not seen major innovations. Today, the current economic situation in Puerto Rico has been marked by almost 10 years of recession, where the real estate market has being affected. The real

estate prices have dropped and foreclosures rates have increased. These combinations have created an increase of for sale properties on the market and a lack of buyer or demand for the market of real estate. This could define the ideal time for the real estate professionals to rely on technology to work smarter on their businesses.

Therefore, this study seek to prove the huge advantage that real estate professionals when applying data mining techniques on data and information available to improve services and/or develop business strategies to improve profitability. The majority of these professionals don't use available data to gain competitive advantage and identify areas where they could improve their services.

METHODOLOGY

This research implemented data mining techniques using public databases for the Home Mortgage Disclosure Act (HDMA) records from 2007 until 2017 to prove the valuable benefits that BI brings to businesses.

The HMDA records have 187,462,446 records from the 50 states of the United States, Washington DC and Puerto Rico. The data set extracted for our analysis is all the records that correspond to Puerto Rico and it is compose of 958,975 records from 2007 to 2017.
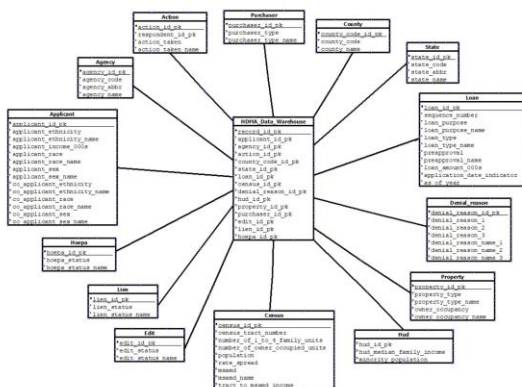

Figure 4
Data Warehouse Composition

As illustrated on Figure 4, this public data warehouse selected to try to identify trends on the

Real Estate transactions consist on information from the US Census, Housing and Urban Development (HUD) and Loan Application Register (LAR). The information from the LAR is provided by the financial institutions and include all the loans transactions from all the financial institutions that had to comply with Regulation C. These institutions are the ones that have over $10 million in assets. Example of the information that this HMDA data set discloses are the following:

- Loan information: The data include mortgage application, type of loan, loan amount, purpose of loan, approval status (approved or denied), and reasons of approval status.
- Demographic information on applicants: The data include ethnicity, race, sex, and income.
- Real Estate Information: The data include type of property, whether the owner intends to live the property, census tract, among other information.
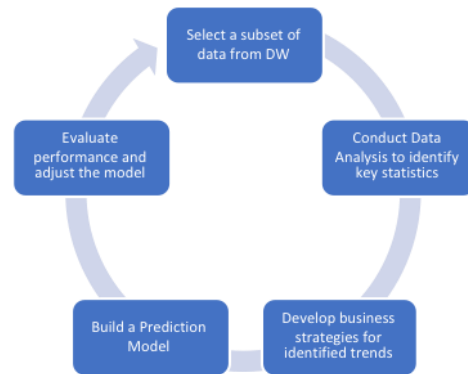

Figure 5
Data Mining Process Used for this Study

This study used the data mining procedure described above in Figure 5. Using the data mining techniques, I will identify trends on the Real Estate transactions that could help the real estate professional to improve the quality of their services that could translate to competitive advantages and effective strategies.

I will use the visualization tool on WEKA to get some insights on trends or information. Although this tool provide some graphics views, I choose to make the data graphics to have a better

quality and clear understanding of the illustration in this paper.

The last step of this study was the development of an initial prediction model for the loan amounts as a parameter of future demand on the real estate industry. For this prediction model, we will use a train data from 2007 until 2016 and then test the accuracy with the actual data from 2017. I will be using the tool of Weka to identify the best prediction model and I will also write a basic machine learning code in Python to analyze the prediction of other models.

RESULTS

After understanding the data, 67% of the records or a total of 647,324 records were used for the Business Intelligence analysis. Our Business Intelligence analysis main objective is to translate the data set into knowledge and potential business strategies for the real estate professional.

Therefore, we continued the analysis of the data and determine the key statistics or valuable information that could be extracted from these data. The data warehouse had a total of 78 different attributes analyzed and normalized. These attributes were composed on numerical and nominal data, and we minimize the attributes identifying the columns that provide the same data on a different format. After minimizing the attributes, the key attributes for the industry identified were:

- Loan Amount Sum per Year – This statistic will reflect the market trends, in terms of supply and demand.
- Loan Amount Distribution – This attribute will provide the range of loan amount with higher demand on the market.
- Loan Originated vs. Denied Applications – This comparison could be interesting because it could be directly related to the quality of consulting services that the real estate agent is providing to the property buyer.
- Reasons for Loan Denial – This attribute could serve the real estate agent for customer service

improvement. Knowing were denial reasons could be an opportunity to help buyers before applying to loans.
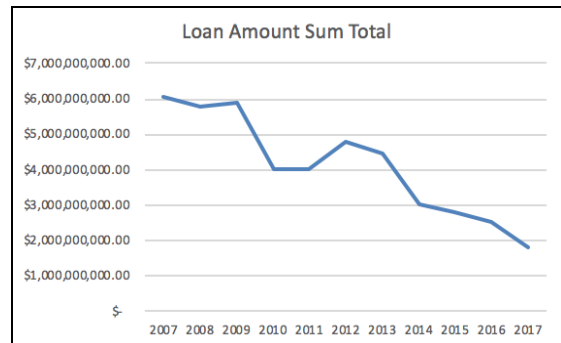

Figure 6
Total Loan Originated Amount per Year

Considering the loans originated from 2007 until 2017, the chart on Figure 6 shows a clear decline trend on loans originated. This trend may reflect the economic situation of recession that the island of Puerto Rico experienced for the past 10 years; that in terms of the real estate industry will mean a low demand of mortgage loans or low amount of property buyers.
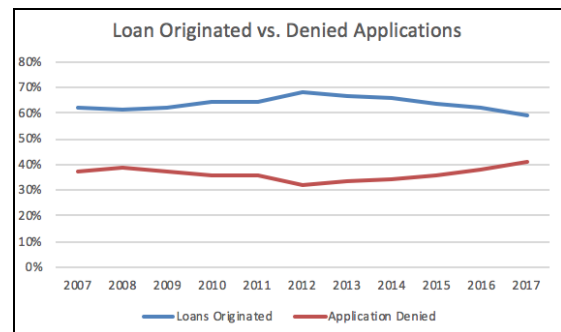

Figure 7
Comparison between Loans Originated and Denied Applications in Percentage

The Figure 7 shows the comparison between the loans originated and the denied applications. This graphics reflect a deflection point on 2012 that will need some additional research to determine the reason for this event. The graph shows an increase on denied applications starting after 2012. The Figure 7 below present the comparison between applications with preapproval requested and not requested. This Figure 8 show an increment of the

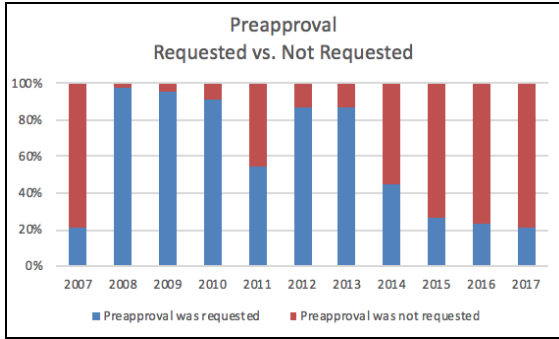percentage of applications that do not requested preapproval.

Figure 8
Comparison between Applications with Preapproval Requested and Not Requested

Analyzing the Figure 7 and Figure 8, there is a direct relation between the increment of denial applications and increment of the applications that do not requested preapproval. Therefore, our hypothesis that stated that buyers follow the Real Estate Professional advise a requested a loan preapproval is contradicted based on the previous analysis. Considering that we are turning data into knowledge to make strategies for the real estate professional, the next step is to look deeper on the denial reasons for these loans.

Table 2
Application's Denial Reasons

DENIAL REASON	%
Credit History	40.8 %
Debt-to-income ratio	23.6 %
Collateral	12.1 %
Other	10.8 %
Insufficient cash (down payment, closing costs)	4.5 %
Credit application incomplete	3.2 %
Employment history	2.7 %
Unverifiable information	2.2 %
Mortgage insurance denied	0.1 %

Table 2 above shows the different denial reasons for the applications on the data analyzed. A 64.4% of the denial reasons concentrate on applicant's credit history and debt-to-income ratio. Therefore, our hypothesis that a low percentage of loans are denied by the debt-to-income because the

Real Estate Professional provide the consultation before the loan application is contradicted.

So, the real estate professional may modify his operating procedures to implement some activities to improve his efficiency and work smarter. Translating the results of this table to strategies, a real estate professional could gain competitive advantage by delivering more consulting services to the buyer of the properties or loan applicants. Considering this strategy, the real estate professional should provide a consulting service to the buyer or the loan applicant considering the following:

- Analysis of the credit history,
- Calculation of the debt-to-income ratio,
- Verification of collateral requirement, and,
- Estimate the down payment and closing cost for a mortgage loan and ensure that the buyer have sufficient cash to cover these costs.
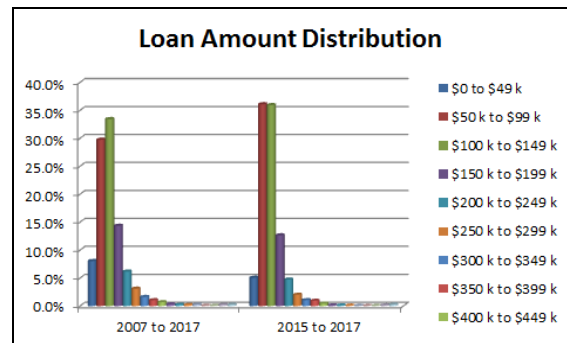

Figure 9
Comparison on the Distribution of the Amount for the Loans between the Complete Data and the Last 3 Years of Data

The above Figure 9 shows the percentage and comparison of the different amounts for the loans applications between the complete data (from 2007 to 2017) and the last 3 years of data. Basically, this data present that more than 72% of the demand of mortgage loans are concentrated on real estate with prices between $50,000 and $150,000 on the last three years. Therefore, a focus on this specific market by the real estate professional could result on more demand and more profit for his work.
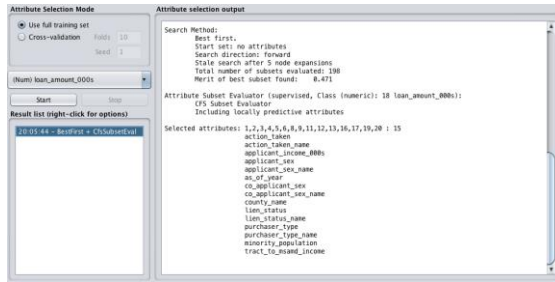
Figure 10
View of the Attributes Selection Output in Weka

Another Business Intelligence technique applied on this study was the development of a prediction model. Using the data of loans originated and denied at Puerto Rico, first we evaluate and selected the best subset of attributes in the dataset that will provide the highest accuracy for the prediction models. Figure 10 shows the result of the attributes selected to predict the loans amount.

After having the selected attributes, the prediction model was developed using the train data of the loans originated from 2007 until 2016. Then these prediction models were evaluated against the actual data for 2017. Table 4 provide the accuracy of the different models evaluated. The accuracy was determined by the mean absolute error using as based the actual data for 2017.

Table 3
Accuracy comparisons of Prediction Models by Regression Algorithm Used

Regression Model Used	Train MAE	Test MAE
Rep Tree	42.51	34.04
XGB Regressor	45	39
Random Forrest Regressor	43	43
Least Median Square	49.36	--
Additive Regression	49.71	--
Iterative Absolute Error	58.79	--

The Rep Tree Regressor was prediction model with lower mean absolute error after the initial tuning. This model build a decision/regression tree using information gain/variances and use reduced-error pruning with back fitting. This model is an included package of the Weka. The other two predictive models that had a close accuracy was created using the python code included as an Appendix of this paper. Considering that the resulting accuracy of the prediction models still high, the model should be improved and should have more tuning to have a more reliable predictions and used it to develop other strategies.

CONCLUSIONS

The World of Business is in constant change and every business that want to survive need to be competitive and gain advantage to be profitable. Despite the long time that BI have been achieving great success histories, different industries have not taken advantage of the benefits from a well-designed Business Intelligence system. Through the implementation of Business Intelligence, business transform historical data into knowledge, improve the decision quality and developed well supported strategies to gain competitive advantage.

As this research present, through the use of data mining techniques, business could translate data into useful information to develop business' strategies to improve services and profitability. The real estate industry have a huge range of data and documents that could provide valuable knowledge. This industry, without a doubt, is poised to be taking advantages of Business Intelligence initiatives in various areas. The future of the real estate professionals will be extremely impacted by the benefits of Business Intelligence.

Today, the real estate professional have enormous opportunities to use the BI and Machine Learning tools available to develop effective strategies and work smarter.

REFERENCES

[1] H. Watson and B. Wixom, "The Current State of Business Intelligence", in *Computer*, September 2007, pp. 96-99.

[2] J. Ranjan, "Business Intelligence: Concepts, Components, Techniques and Benefits", in *Journal of Theoretical and Applied Information Technology*, Institute of Management Technology, Ghaziabad, Uttar Pradesh, India, 2005.

[3] V. H. Trieu, "Getting Value from Business Intelligence Systems: A Review and Research Agenda," in *Decision Support Systems*, University of Melbourne, October 2016.

[4] E. Poulis and A. Thamir, "Business Intelligence Capabilities and Implementation Strategies", in

International Journal of Global Business, University of Manchester, UK, 2015.

[5] S. Chaudhuri, U. Dayal, and V. Narasayya. (2011, Aug.). "An Overview of Business Intelligence Technology", in *Communications of the ACM* [Online]. Available: https://cacm.acm.org/magazines/2011/8/114953-an-over view-of-business-intelligence-technology/fulltext.

[6] R. Agarwal and V. Dhar. (2014, September 25). *Editorial – Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research* [Online]. Available: https://pubsonline.informs.org/doi/full/10.1287/isre.2014.0 546.

[7] T. Loughhead. (2014, December). *What the rise of machine learning means for real estate sales* [Online]. Available: https://www.inman.com/2016/12/14/rise-machi ne-learning-means-real-estate-sales/.

[8] S. Adelman, L. Moss and L. Barbusinski. (2002, *August 17). I found several definitions of BI, DM Review* [Online]. Available: http://www.dmreview.com/article_sub.cfm?arti cle=5700.

[9] S. Negash, "Business Intelligence", in *Communications of the Association for Information Systems*, vol. 13, 2004.

[10] D. Farmer and M. Kohtamaki, "Strategic Agility – Integrating Business Intelligence with Strategy", in *Real-time Strategy and Business Intelligence*, 2017.

[11] J. Hoffmeister. (2017, November 13). *The growing importance of machine learning in real estate transactions* [Online]. Available: https://www.itproportal.com/fea tures/the-growing-importance-of-machine-learning-in-real-estate-transactions/.

[12] B. Berger. (2017, September 19). *AI is changing how real estate professionals do business* [Online]. Available: https://venturebeat.com/2017/09/19/ai-is-changing-how-real-estate-professionals-do-business/.

[13] M. Krigsman. (2017, July 30). *Zillow: Machine learning and data disrupt real estate, Beyond IT Failure* [Online]. Available: https://www.zdnet.com/article/zillow-machine-learning-and-data-in-real-estate/.

[14] D. Harris. (2017, July 21). *How Does Machine Learning Work?* [Online]. Available: https://www.articlecity.com/ blog/machine-learning-work/.

[15] G. Levy-Yurista. (2018, February 27). *The Next Generation of BI is here… and so is the 2018 Gartner Magic Quadrant for analytics and Business Intelligence Platforms* [Online]. Available: https://www.sisense.com/ blog/gartners-2018-magic-quadrant-next-generation-bi/.