Master of Science in Engineering

Success Drivers of Online Real Estate Crowdfunding Using Platform Data

by

Perry Whitecage

Department of Architecture & Architectural Engineering

The Graduate School

Seoul National University

February 2018

Success Drivers of Online Real Estate Crowdfunding Using Platform Data

by

Perry Whitecage

A thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Engineering

Seoul National University

2018

Success Drivers of Online Real Estate Crowdfunding Using Platform Data

December, 2017

Approved by Dissertation Committee:

Hyun-Soo Lee

Moonseo Park

Seokho Chi

Success Drivers of Online Real Estate Crowdfunding Using Platform Data

지도교수 박 문 서

이 논문을 공학석사 학위논문으로 제출함

2017년 12월

서울대학교 대학원

건 축 학 과

페리

페리의 공학석사 학위논문을 인준함

2017년 12월

위 원 장 이 현 수 (인)

부위원장 박 문 서 (인)

위 원 지 석 호 (인)

Abstract

Success Drivers of Online Real Estate Crowdfunding Using Platform Data

Perry Whitecage
Department of Architecture and Architectural Engineering
Graduate School
Seoul National University

Real estate crowdfunding, the raising of a number of relatively small amounts of capital from a large number of people (the "crowd"), has gained widespread popularity in recent years and has the potential to provide financing for an increasing share of real estate. While real estate crowdfunding is the fastest growing segment of the global crowdfunding industry, stakeholders have little guidance on what are the success drivers which motivate backer's investment decisions.

The purpose of this exploratory study is to gain insight into the relevant factors that influence both funding success and the amount of days it takes a solicitation to meet or exceed its target commitment amount based on the data provided to potential investors on the online crowdfunding platforms. This research utilizes an open source preprocessing tool and machine learning algorithm collection in a multiple step knowledge discovery process. The dataset consists of 275 debt offerings with 16 attributes from a leading real estate crowdfunding platform in the United States.

This study is the first to use data mining of platform data to explore the success drivers for online real estate crowdfunding, providing owners, developers, managers and the crowdfunding platforms with insights that can support the decision to use crowdfunding and how to design projects and offerings for funding success. Results reveal the subset of factors which are most and least relevant to motivating backers and indicate that the factors which are relevant differ between residential and commercial real estate offerings. Findings also reveal that the criteria for motivating backers in a crowdfunding context are different from other real estate investments and that real estate crowdfunding has some similarities and differences compared to equity and reward crowdfunding.

Keywords: Real Estate Crowdfunding, Investments, Title II, Online Platforms, Data Mining

Student Number: 2016-22085

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background

Real estate crowdfunding, the raising of a number of relatively small amounts of capital from a large number of people (the "crowd"), has gained widespread popularity in recent years and is expected to grow exponentially, thus providing financing for an increasing share of real estate (Massolution Industry Report, 2015). Figure 1.1 shows the global growth of real estate crowdfunding up to 2015. In addition, real estate crowdfunding is forecasted to continue to grow to a total of $250 billion by 2020 (Massolution Industry Report, 2015). Individual campaigns can range in size from less than $100,000 to over $25 million.

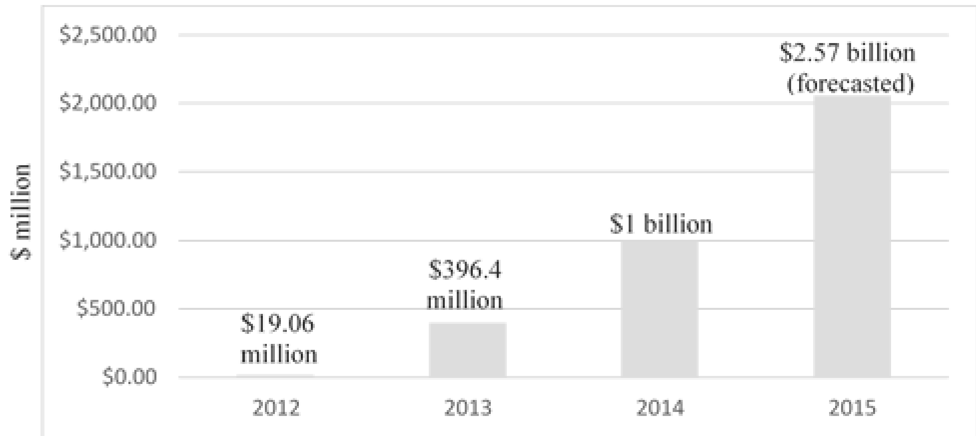

Figure 1.1 – Growth of Global Real Estate Crowdfunding

As a relatively new and unique method of funding, real estate crowdfunding has different characteristics from other real estate investments, and there is little prior data-driven research on the characteristics of real estate projects that succeed using this alternative form of finance. While real estate crowdfunding is the fastest growing segment of the global crowdfunding industry, stakeholders have little guidance on what are the success drivers which motivate backer's investment decisions.

While there have been many successfully crowdfunded real estate transactions, the sheer size and relevance of the marketplace is still relatively immature. As a result, there is little guidance on the characteristics of the projects which are suitable for crowdfunding. It has become clear that to advance, firms must make industry-defining choices about what size, type of asset, and capital structure to pursue (Cohen, 2016).

1.2 Problem Statement & Research Objectives

Crowdfunding of real estate has grown rapidly, yet there is little research on the factors that motivate the relatively inexperienced investors participating in real estate crowdfunding. The real estate crowdfunding platforms provide information online from which these investors evaluate the quality of the solicitation. Adjacent crowdfunding research has yielded significant results using a data mining approach for gathering insights into the relevant factors that determine crowdfunding success; however the research focused on real estate crowdfunding is limited. This research aims to explore which factors from the information provided on the crowdfunding platforms by the sponsors of the projects are relevant to motivate investors to fund the project.

The purpose of this exploratory research is to gain insight into the relative importance of different drivers (influences) on funding success and the amount of time it takes for a solicitation to succeed from the information provided on the platforms. The potential drivers will be selected from the platform data. A knowledge discovery process utilizing data mining methods will be conducted on that data to yield insights on the success drivers.

1.3 Scope of Research

The research data is from a leading real estate platform that is limited to debt (loans) for real estate within the United States. The scope of this research is limited to US online 506(c) offerings, as regulated by Title II of the Jumpstart Our Business Startups (JOBS Act), which was passed in 2012. Title II limits funding commitments to accredited investors, which include individuals with income in excess of $200,000 per year in the last two years ($300,000 combined income if married) or net worth over $1 million (excluding their primary residence). This study is limited to data mining only a set of attributes from the platform data that is provided to potential investors. The dataset excludes data in any attachments, such as property appraisals, or data found on linked websites which are external to the crowdfunding platform.

1.4 Research Methodology

The research process can be defined in the three steps as shown in Figure 1.2. The first step is to explore the problem and develop the research questions, review the literature and study the potential variables. The second step is research design consisting of the data-mining methods, developing the theoretical background, and developing a model. The last step is research execution where we collect the data, preprocess the data, conduct data analysis and draw conclusions.

This research explores the data utilizing WEKA (Waikato Environment for Knowledge Analysis), which is a data-preprocessing tool and machine learning algorithm collection for data mining provided by the University of Waikato in New Zealand.

The thesis is structured as follows. We start with a definition of crowdfunding followed by the theoretical framework from which the study is based. Next is a look at how traditional real estate and how crowdfunding ventures have separately been evaluated. Following that is a section on knowledge discovery in databases and data mining. Next we explore the variables used in this research and develop a model. The penultimate section describes the data, the knowledge discovery process in detail, the results of the analysis and a discussion of the findings. We conclude with a summary, limitations and some recommendations for further research.

This research would be relevant in two aspects. The first is the scientific contribution to the theory of knowledge discovery for engineering, construction & real estate industry using data-mining techniques. The second is a practical contribution of a method to provide decision support for owners, developers, managers and platforms on whether to use the crowdfunding model to raise funds and how to design projects and campaigns for funding success.

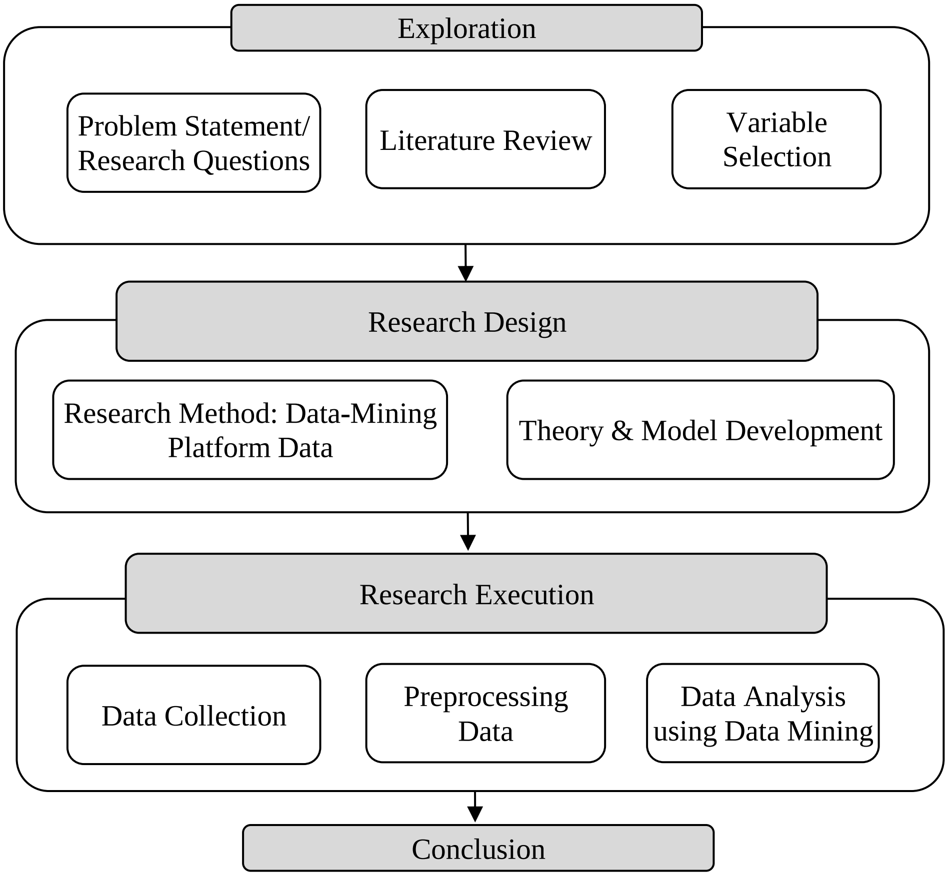

Figure 1.2 – Research Process

Chapter 2. Preliminary Study

This chapter begins with the definition of real estate crowdfunding, followed by the theoretical framework for this study, and how real estate crowdfunding compares to other real estate investment options. It also reviews the prior research about evaluating real estate, evaluating crowdfunding ventures, and using a data mining approach.

2.1 Crowdfunding

Online crowdfunding is a relatively new method of raising money from a large group of ordinary people, in which each individual provides a small amount, rather than raising a large amount from a small group of typically sophisticated investors. This alternative financing method leads to new forms of business development in which the "ordinary" crowd gets more closely involved, as active consumers, investors, or both (Bellaflame et al., 2012).

Crowdfunding can also be thought of as the junction of crowdsourcing and finance. Crowdsourcing has a number of definitions, but in common is the idea that it invites all interested people to form an open forum of ideas that can eventually lead to a solution of the assigned problem (Misra et al., 2014).

By raising money online, crowdfunding has a different stakeholder environment from traditional financing methods. A number of prior studies have defined the stakeholders in crowdfunding and their roles. Tomczak &

Brem (2013) identify three stakeholders of crowdfunding: entrepreneurs, investors and intermediary. Valančienė and Jegelevičiūtė (2014) also identify three stakeholders, referring to them as businesses, backers and platform. In common is the general process of funding using the online platform as an intermediary between the business and the backer/ investor.

The online platforms allows businesses to present their ideas or ventures for the general public and solicit funding. The crowdfunding platforms publicize these ideas or ventures, creating an investment possibility for ordinary people. These potential backers analyze the proposed ideas and choose the ones they believe in to fund. As backers like and believe in the funded project, and desire for it to succeed, they tend to (if a possibility) provide advice for the business (Valančienė and Jegelevičiūtė, 2014). Businesses then offer the backers something in return for their money that acts as a reward, such as a small gift, equity, interest (for debt), or a percentage of revenue. When an idea, venture or project is successfully crowdfunded, the businesses usually are obligated to pay a fee to the online platform. Figure 2.1 shows the crowdfunding stakeholders and how they are linked (Valančienė and Jegelevičiūtė, 2014).

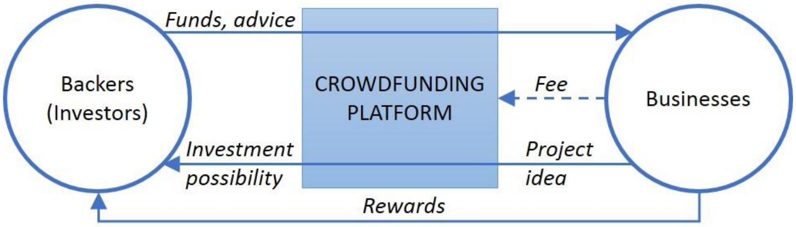

Figure 2.1 - Crowdfunding Stakeholders and How They are Linked

Real Estate Crowdfunding enables investors to access pre-vetted real estate investment opportunities and invest passively in real estate (Schweizer & Zhou 2016). The business (a developer or other entity who sponsors the project) submits proposals to the platform (the intermediary) who does background checks on the projects. In practice, the crowdfunding websites emphasize their ability to find outstanding sponsors and perform careful due diligence on these sponsors (Vogel & Moll 2014). If a sponsor makes it through the background check and screening process, they are allowed to present deals on the platform.

Investors (the backers) can choose either debt or equity investments. Debt investments are typically loans that are tied to a specific property, and secured by it until repaid. Equity investments are usually made by purchasing shares in a limited liability company LLC that invests in a limited partnership (LP) that holds the property.

As a unique method of funding, real estate crowdfunding has different characteristics from other real estate investments. Directly held real estate and pre-sales are transparent to investors in that they are participating in a specific property and they have control of the investment. Similar to crowdfunding, REITS (Real Estate Investment Trusts) provide low minimum investment, low transactions costs, and proportional ownership. However, REITS provide less transparency and control compared to directly held real estate and pre-sales. Real Estate crowdfunding provides the benefits of REITS while maintaining the transparency and control of the investment provided by directly held real estate and pre-sales. Table 2.1 compares real estate crowdfunding to other real estate investment options.

Table 2.1 - Comparison of Real Estate Investment Options

	Low Minimum Investment	**Low Transaction Costs**	**Proportional Ownership**	**Transparent Investment**	**Control of Investment**
Directly Held Real Estate				✓	✓
Pre-Sale				✓	✓
REITS	✓	✓	✓		
Real Estate Crowd-Funding	✓	✓	✓	✓	✓

Real estate crowdfunding is also different in the makeup of the investors. Manonov et al. (2017) found clear differences between a syndicate-based model (traditional model) and a crowdfunding model. Investors in real estate crowdfunding tend to be relatively inexperienced and unsophisticated. Compared to institutional investors, “the crowd” of individual investors, who are often the primary target of project developers on real estate crowdfunding platforms, do not normally have the ability to research or assess such investments (Ahlers et al., 2015). In this light, the information that is provided by the platforms takes on increased significance.

2.2 Communication from Borrowers and Signaling Theory in a Real Estate Crowdfunding Context

In order for a solicitation to successfully get funded via a real estate crowdfunding platform, the sponsors of the campaign (the borrowers) need to clearly communicate their value to investors, who must possess a certain amount of judgment to make informed decisions (Schweizer & Zhou, 2016). Adjacent research in equity and reward-based crowdfunding is consistent with the view that potential backers try to evaluate venture quality by interpreting information from the platforms (Ahlers et al., 2015, Manonov et al., 2017).

The theoretical background for the current research is signaling theory from management literature. This theory is useful for describing behavior when two parties (individuals or organizations) have access to different information (Connelly et al., 2011). Typically, one party, the sender, must choose whether and how to communicate (or signal) that information, and the other party, the receiver, must choose how to interpret the signal (Connelly et al., 2011). For real estate crowdfunding, the platform data is the signal from which the investors evaluate the offering. Figure 2.2, shows the signaling timeline within a crowdfunding context adapted from Connelly et al. (2011).

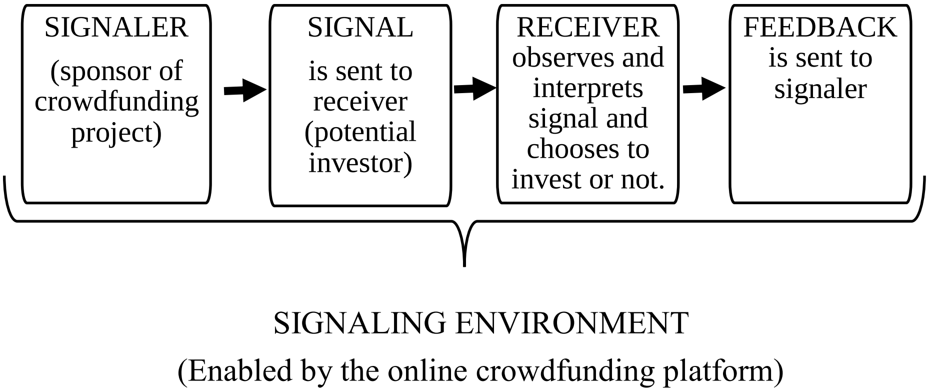

Figure 2.2 - Signaling Timeline within a Crowdfunding Context

2.3 Evaluating Real Estate in Traditional Context

Methods for evaluating real estate have been an important field of research for the real estate and construction industry. A number of prior researches have looked at the relevant factors that are used by sophisticated investors when evaluating real estate in the traditional context. These efforts have been done to establish the important factors relevant to gauging quality of a property and provide a starting point for the current research regarding variables and how they interact.

The objective of Fisher et al. (2004) is to identify the relative correlation of market, owner and property-specific variables, with the likelihood of investment-grade property sales activity. Market factors include four variables; economic, demographic, financial and taxation. Owner factors include organizational and operational characteristics. Property factors include the condition and age of structure, location, type, and viability of tenants. The study researches whether the relative importance of these different factors varies across different types of property. They conclude that three factors (market, owner, property) play significant, independent and equivalent roles in the probability of a sale (Fisher et al., 2004).

A United States patent by Mahlon Apgar, IV (1997) has been filed with the purpose to provide real estate evaluations in an objective, cost-effective, timely and quantitative manner. The invention provides decision support to identify essential factors driving real estate decisions. Information is processed to determine indicators of Amount, Price, Area, Grade, and Risk. These indicators are combined to provide a total score and after the processing, a report is generated which has the score and detailed information, to provide an overall picture of a specific real estate situation. Table 2.2 shows the details of the five variables.

Table 2.2 - Variables used in U.S. Patent by Mahlon Apgar, IV (1997)

Category	Metric
Amount	Space utilization of real estate (sq. ft. per employee, and/or sales or revenues per sq. ft.).
Price	Cost utilization of real estate (rent per sq. ft and/or rent per employee, and/or rent to sales).
Area	Economic attractiveness of submarket location (rents, vacancy, absorption rate, etc.).
Grade	Quality of real estate (Class A, B or C properties).
Risk	Exposure of real estate to financial, market and environmental risks.

2.4 Evaluating Crowdfunding Ventures

Prior researchers have also studied how backers evaluate crowdfunding ventures. Ahlers et al. (2015) building on the work of Baum & Silverman (2004) develop a framework that describes the connection between venture quality and uncertainty (independent variables) on fundraising success (dependent variable). Specifically, they look at 104 offerings, between October 2006 and October 2011, from the Australian Small Scale Offerings Board (ASSOB) which at the time was one of the largest equity crowdfunding platforms. As shown in figure 2.3, the model is that venture quality positively contributes to the probability of funding success and the level of uncertainty negatively contributes to the probability of funding success (Ahlers et al., 2015).

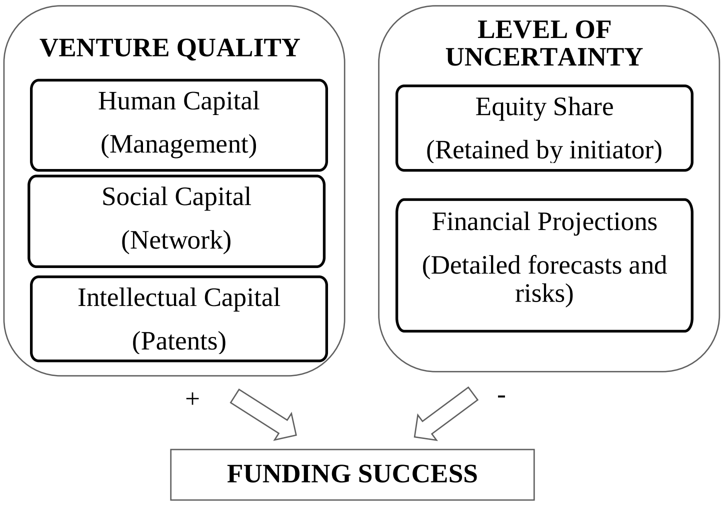

Figure 2.3 - Model of Funding Success

Additional researchers have also attempted to find the influential factors that correlate with success. Manonov et al. (2017) finds that research across multiple platforms indicate the size of the requested funding is generally negatively correlated with funding success.

However, other research indicates financial factors are not the most important. Lukkarinen et al. (2016) find that investment decision criteria traditionally used by VCs or business angels are not of prime importance for success in equity crowdfunding; success is related to pre-selected crowdfunding campaign characteristics and the utilization of private and public networks. Lukkarinen et al. (2016) also find that emotional and social criteria may be more important to backers than financials.

2.5 Process of Knowledge Discovery in Databases

The current research methodology is generally referred to as knowledge discovery in databases (KDD). The KDD process consists of several methods (Silwattananusarn & Tuamsuk, 2012, Fayyad, et al., 1996).

1. Selection: Selecting data relevant to the analysis task from the database.
2. Preprocessing: Removing noise and inconsistent data; combining multiple data sources.
3. Transformation: Transforming data into appropriate forms to perform data mining.
4. Data mining: Choosing a data mining algorithm which is appropriate to pattern in the data; extracting data patterns.
5. Interpretation/Evaluation: Interpreting the patterns into knowledge by removing redundant or irrelevant patterns; translating the useful patterns into terms that are human understandable.

Figure 2.4, sourced from Fayyad et al. (1996) shows the iterative sequence.

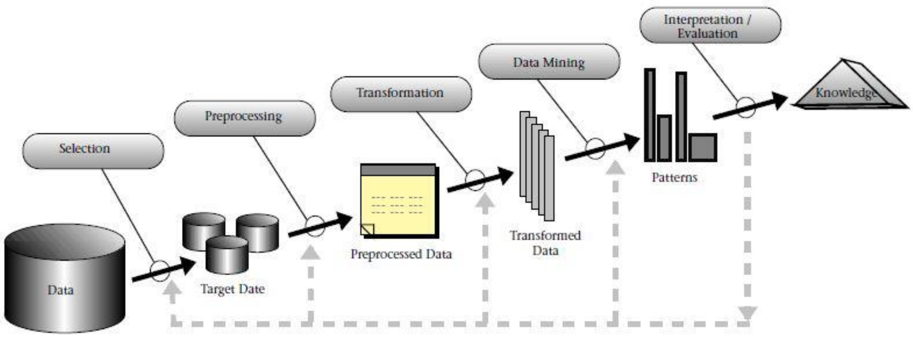

Figure 2.4 - Process of Knowledge Discovery in Databases and Data Mining

2.6 Prior Research Using Data Mining for Insights into Backer Motivation in Crowdfunding

There have been a number of studies into backer motivation for crowdfunding ventures using a data-mining approach. Data mining is the application of specific algorithms for extracting patterns from data and is part of a more general knowledge discovery in databases (KDD) process (Fayyad et al., 1996). KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et al., 1996).

Rakesh et al. (2015) have the objective to determine what set of features determine a Kickstarter project's success. They used a supervised learning framework to create a model that can recommend potential backers with results that are significantly better than the previous studies. They set-up the framework as a binary classification problem. The trained model computes the score that represents the likelihood of funding after being given a backer-project pair.

Yuan et al. (2016) conducted research to identify the most influential topical features embedded in project descriptions to better promote projects and improve funding success. They use a semantic text analytics approach and focus on crowdfunding dynamics in China. The proposed framework outperforms a classical method in predicting the success of funding by an average of 11% (Yuan et al., 2016).

Most applicable to the current research is Manonov et al. (2017), whose objective was to understand how Title II crowdfunding fits into the larger crowdfunding landscape and the types of ventures that have succeeded using crowdfunding. They further explored factors into why residential and commercial real estate had the highest success of all industries using Title II crowdfunding. However, they were limited to text mining of project descriptions. Their process consisted of a bag of words transformation of project descriptions to create a feature set and then they used naïve Bayes classification on 388 real estate offerings. Within that scope, they were able to build accurate models predicting success from a series of lexical indicators. Similar to others, they used a binary classification model; success or failure.

There are two main points from Manonov et al. (2015) most relevant to our research. First, there are clear differences between a syndicate-based model (traditional model) and a crowdfunding model. The syndicate model relies on a community of venture capitalists or others to perform the necessary task of due diligence, screening, and selection (Manonov et al., 2015). In contrast, those functions and processes are done by the platform in the crowdfunding model. Second, the text mining of project descriptions reveals the importance to investors of the platforms performing due diligence on potential opportunities. However, it is the limitation of only using text mining which is the knowledge gap our research is attempting to fill.

2.7 Summary

Crowdfunding has a unique stakeholder environment that involves three primary stakeholders. Signaling theory is the theoretical background of this research and supports the validity of using the online crowdfunding platform data as the means by which investors make decisions about venture quality. The role of the platform as the intermediary between the backer and the sponsor creates a signaling environment in which the platform data is used to evaluate the quality of the offerings and motivate the backers to fund the offerings. Real estate crowdfunding is different from other real estate investment options and the relatively inexperienced investors are different from investors in the traditional context. The prior research about evaluating real estate and evaluating crowdfunding ventures provides us a starting point for looking at variables and how they interact. And finally, prior research into using a data mining approach gives a direction on using data mining, insights into the limitations of that research and the knowledge gap to be filled.

Chapter 3. Variable Selection of Success Drivers & Model of Backer Motivation

This chapter explores the variables that can be extracted from the crowdfunding platform data. It also develops a model for the current research.

3.1 Variables from Platform Data

The variables for this research are derived from the information provided by the platforms and are used as the attributes that are analyzed as drivers of funding success. The information provided is the signal from the sponsor and the platform that has the potential to motivate the investor to fund the solicitation.

Within real estate crowdfunding, there can be different services that any given platform can perform, and therefore there will be different information provided. However, there are several unifying components of real estate crowdfunding that can be seen within the majority of platforms.

Platforms often include a value proposition in which the nature and configuration of the site are clearly laid out along with other fundamental features that must be addressed (Cohen 2016). It is these features that comprise the variables of the current research. The features include the types of properties, investment structure, holding periods and other information from which investors can evaluate the value proposition (Cohen 2016).

For this research, we separated the features into property, borrower and financial factors. The dependent variable, representing the motivation of the investors, is the total number of days to fund the loan. Figure 3.1 shows the variables that can be extracted from the platform data. The complete description of these variables can be found in chapter 4.

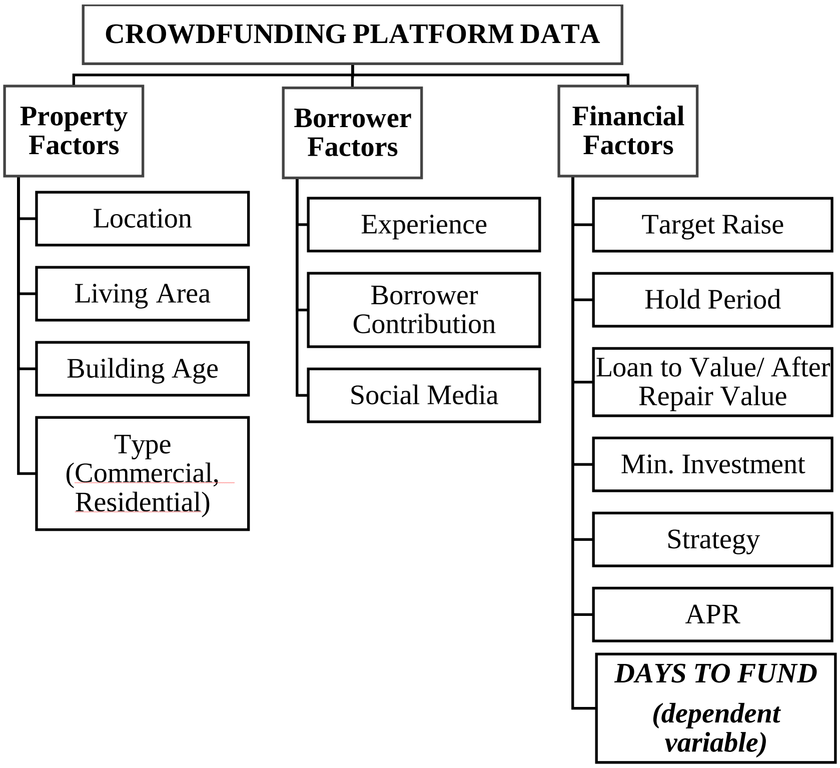

Figure 3.1 – Variables from Platform Data

3.2 Model of Backer Motivation

This research aims to build a model of the factors extracted from the information provided on the crowdfunding platforms by the sponsors of the projects. The completed model should distinguish between the factors that are important and unimportant to motivate investors to fund the project. Furthermore, the factors that are important should be distinguished between those that positively affect motivation and negatively affect motivation of the backers. Figure 3.2 shows the schema of the research model.

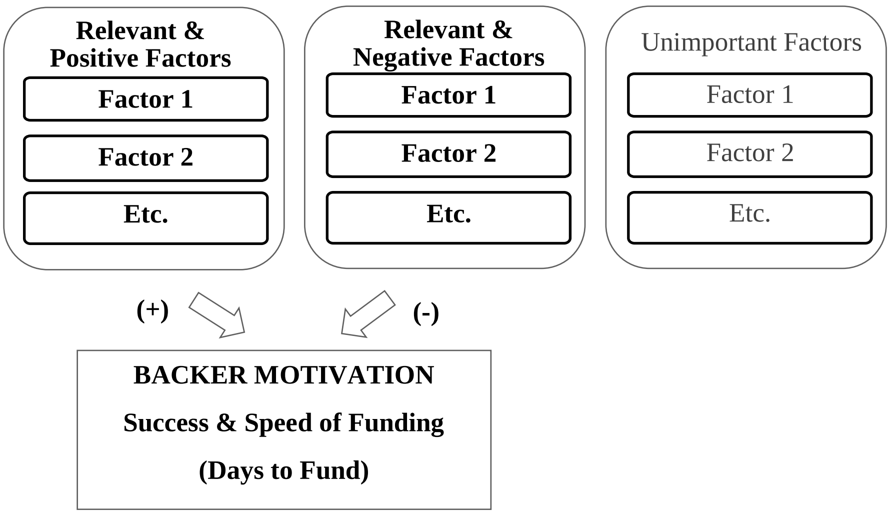

Figure 3.2 –Schema of the Research Model

3.3 Summary

The variables that can be extracted from the crowdfunding platform data can be separated into three categories; property, borrower and financial factors. The model for the current research is that there will be relevant factors that affect the motivation of the investors in positive and negative ways, and there will also be factors that are unimportant.

Chapter 4. Analysis of Crowdfunding Platform Data

This chapter discusses the specific online crowdfunding platform and the data used in this research and how the dataset was preprocessed and analyzed, followed by the results of the analysis and a discussion of the findings.

4.1 Platform & Data

All projects are from the online crowdfunding platform found online at www.patchofland.com (POL), operated by Patch of Land Lending, LLC which is a wholly-owned subsidiary of Patch of Land, Inc. POL is headquartered in Los Angeles, California and was founded 2013. Table 4.1 shows some statistics about Patch of Land.

Table 4.1 – Patch of Land Statistics

655	Total Successful Loans Funded
66.41%	Weighted Average Loan to Value at initial loan disbursement (since April 2015)
11.12%	Realized Rate of Return
$487,642	Average Loan Size Total Funds Returned to Investors
$93,165,642	Total Funds Returned to Investors
$319,446,952	Total Loans Funds (Through 2nd quarter, 2017)

The initial (raw) dataset contains 296 closed records, with 23 attributes, from offerings made between March 2015 and August 2017. The dataset for this study was obtained from FinMkt, a New York City-based crowdfinancing firm founded in 2011. 25 offerings failed to reach target funding, 102 reached their target funding, and 169 exceeded their target funding. See Appendix A for a complete list of the original attributes and their descriptions.

4.2 Knowledge Discovery Process & Data Mining

The process of getting useful information from data mining involves several steps, with the algorithms used just one part (Silwattananusarn & Tuamsuk, 2012). First, we collected the data which consists of the original dataset from FinMkt and then we manually added some information that is prominent on the POL platform but was not in the raw dataset, such as APR and the size of the properties.

After data collection, we performed extensive data pre-processing on the entire data-set. Next, we split the data-set into residential and commercial properties, providing us with two separate data sets to analyze. Then we conducted feature selection using a correlation algorithm on the data-sets to find the relevant attributes. Finally, we performed data clustering, an unsupervised data mining technique, used to find similarity in groups of data. Figure 4.1 shows the overall knowledge discovery process for this research.

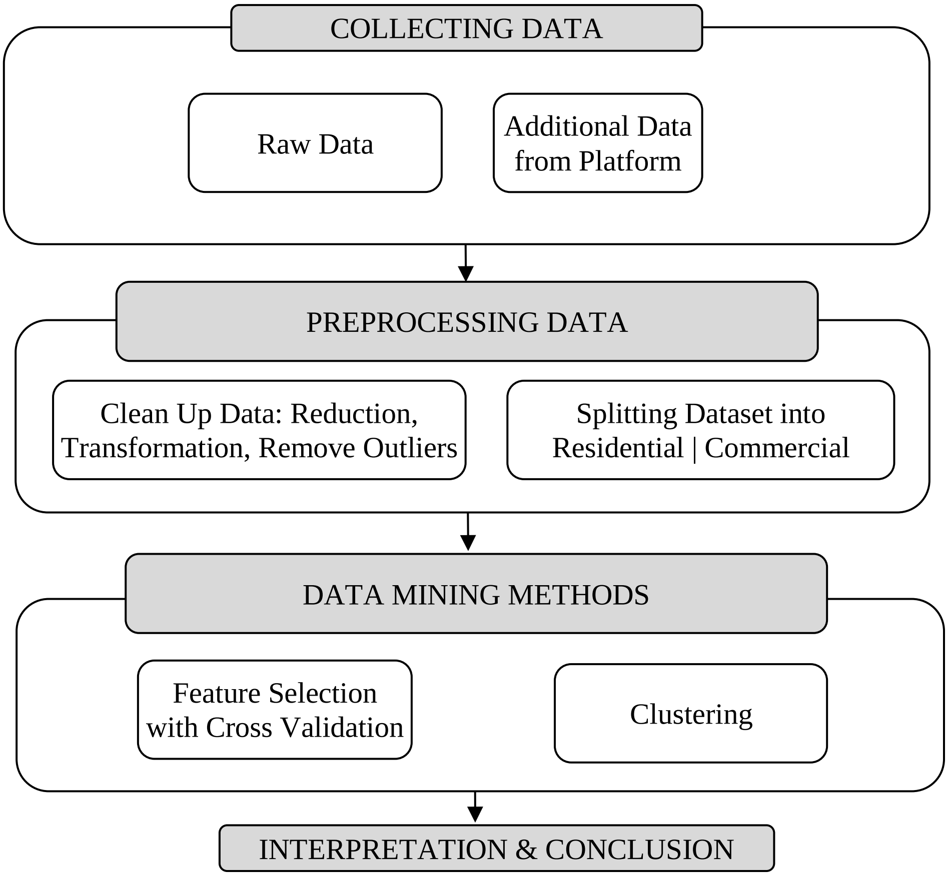

Figure 4.1 – Knowledge Discovery Process of Current Research

4.3 Preprocessing Platform Data

The first preprocessing step is data reduction which involved two methods to reduce data dimensionality. The first method is to remove unnecessary attributes which are attributes that have the same value for all records. For example, the attribute “Sector” has value “real estate” for all records so can be removed because it adds no information. The second method of dimensionality reduction is to aggregate attributes that can be combined without loss of information. For example, original data had four attributes for Social Media (LinkedIn, Facebook, Twitter and GooglePlus), but all businesses either had all four or had none, thus the attributes could be aggregated into a single binary attribute with value of “yes,” or “no”. After reduction & aggregation, there were 16 attributes remaining.

The next step in preprocessing is transformation which is a process that converts the data to be more useful for data-mining. For example, “Year Built” (1950) would be converted to “Age” (77).

The next preprocessing step is removing outliers or extreme values. Using Weka’s interquartile method, 21 outliers were identified and removed, leaving 275 records remaining for analysis. See appendix B for the details of all the data preprocessing steps. Table 4.2 has the description of the dataset’s 16 attributes used for analysis.

Table 4.2 – Attribute Details

Category	Attribute	Unit	Detailed Description
Property Factors	City	Nominal	City where the Company's headquarters or principal place of business is located.
	State	Nominal	State where the Company's headquarters or principal place of business is located.
	Gross_Living_ Area	Sq. Ft.	Gross Living Area in square feet of subject property, not including the area of the lot.
	Building_Age	Years	Age in years of subject property.
	Property_Type	Nominal	The type of property to which investments are directed, and may include, though not be limited to, industrial, commercial or residential.
Borrower Factors	Borrower_ Experience	Years	Years of borrower experience as extracted from borrower description.
	Borrower_ Contribution	USD	Financial contribution of the borrower for renovation or completion depending on strategy.
	Contribution_ Ratio	Percentage	Financial contribution of the borrower as a ratio of the target raise that the borrower is seeking.
	Social_Media	Nominal	A binary value (yes/ no) indicating social media urls provided by the borrower for LinkedIn, Facebook, Twitter and GooglePlus.
Financial Factors	APR	Percentage	An annual percentage rate (APR) is the annual rate charged for borrowing.
	LTV_or_ARV_ Ratio	Percentage	Loan to Value Ratio (LTV) represents an amount borrowed as compared to the value of the property. After Repair Value Ratio (ARV) represents an amount borrowed as compared to an estimated value of a property after renovations.
	Hold_Period	Months	This is the anticipated amount of time that a property is held before exiting the investment.
	Min_Investment	Nominal	Minimum Investment Amount as binary value where “yes” indicates 5000 USD minimum.
	Target_Raise	USD	Total amount of the offering.
	Strategy	Nominal	This is the intended strategy to add value to an investment.
	Days_to_Fund	Days	Amount of days to reach or exceed target raise with 0 indicating failure to raise target amount.

4.4 Splitting Data into Residential and Commercial Datasets

The final step before analysis is splitting the dataset into two separate datasets; one containing only single family residential properties and one for commercial properties. The purpose of this step is to find the relevant success drivers that apply to these distinct investment groups and compare the insights.

The results were a dataset of 190 records for single family residential and a dataset of 79 records for commercial. Seven records for condominiums were excluded from both sets because of the uniqueness of the property type and the small sample size.

4.5 Selection of Relevant Features Using Pearson′s Correlation Coefficient

A correlation technique is used to select the most relevant attributes in the research datasets. WEKA has correlation based feature selection using the CorrelationAttributeEval technique. This technique calculates the correlation between each attribute and the output variable (Days_to_Fund).

In statistics this correlation is referred as Pearson′s correlation coefficient which is the covariance of the two variables divided by the product of their standard deviations. The formula for Pearson's correlation coefficient is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

r = [Equation]
Σ = Number of pairs of sources
Σxy = Sum of the products of paired scores
Σx = Sum of x scores
Σy = Sum of y scores
Σx^2 = Sum of squared x scores
Σy^2 = Sum of squared y scores

Pearson's correlation coefficient measures the strength of the association between two variables and attempts to draw a line of best fit through the two variables data points. Pearson's correlation coefficient indicates how well the data points fit this line of best fit. Table 3.3 shows the results of feature selection and Appendixes C and D shows the complete run information.

Table 4.3 – Top Five and Bottom Five Rankings of Feature Selection

Rank	Single Family Residential	Commercial
1	Target_Raise	Target_Raise
2	Hold_Period	APR
3	Strategy	Hold_Period
4	LTV_or_ARV_Ratio	Social_Media
5	Min_Investment	Property_Type
12	Social_Media	LTV_or_ARV_Ratio
13	Borrower_Contribution	State
14	Building_Age	Building_Age
15	Gross_Living_Area	City
16	City	Gross_Living_Area

The findings for single family residential properties are unambiguous. The top five relevant features for single family residential properties are all financial factors. The bottom five are property factors along with two borrower factors (social media and borrower contribution).

The findings for commercial properties are more mixed than for single family residential. The top three relevant features for commercial are financial, and also included one borrower factor (social_media) and one property factor (property_type). Property_type can be an expected attribute in commercial, with a mix of different property types, rather than single family residential which only had one property type. Commercial included hotel, retail, multi-family residential and mixed-use among others. Surprisingly, LTV_or_ARV_Ratio, which is ranked fourth in residential is 12th for commercial. The other least relevant factors are all property factors.

4.6 Segmentation of Data using Clustering

This research uses K-means clustering, an "unsupervised" learning technique conducted in Weka using the Simple K-means algorithm. The goal is to minimize variability within clusters, and maximize variability between clusters to allow for exploration of similarity of members of the group. The K-means algorithm is a classical and well known clustering algorithm and the most commonly used partitioned clustering algorithm because it can be easily implemented and is efficient in terms of the execution time (Ahmad et al., 2015).

The general steps of the K-means algorithm are as follows (Gorunescu, 2011):

1. Select k points at random as cluster centers.
2. Assign instances to their closest cluster center according to some similarity distance function.
3. Calculate the centroid or mean of all instances in each cluster.
4. Cluster the data into k group where k is predefined.
5. GOTO step 3 and continue until the same points are assigned to each cluster in consecutive rounds.

For clustering on the residential dataset we use the Manhattan distance function and random initialization. The following table shows the results of residential clustering.

Table 4.4 – Results of K-means Clustering on Residential Dataset

Attribute	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Target_Raise	173290	278000	109000	810115	295200
Hold_Period	12	12	12	12	12
Strategy	Purchase & Rehab	Refinance & Rehab	Refinance	Refinance & Complete	Purchase
LTV_or_ARV	63	65	60	59	75.5
Min_Investment	Yes	Yes	Yes	Yes	Yes
Days_to_Fund	2	0	1	102	8

It is important to remember that cluster analysis is an exploratory tool and all algorithms have some ambiguity in some (noisy) data when clustered (Ahmad et al., 2015). As a result, findings need to be interpreted.

For residential we can observe that as the target raise goes up, the duration to fund is longer, with less affect from strategy or affect from measure of risk (indicated by fluctuating LTV or ARV). Inferring from Clusters 1 & 4 - the difference between a similar target raise failing or succeeding in a relatively short 8 days is the strategy. The purchase succeeded even though it had higher risk (as measured by LTV).

For clustering on the commercial dataset we use the Manhattan distance function and farthest first initialization. The following table shows the results of commercial clustering.

Table 4.5 – Results of K-means Clustering on Commercial Dataset

Attribute	Cluster 0	Cluster 1	Cluster 2
Target_Raise	177625	175000	553500
APR	11	10	10.9
Hold_Period	12	12	12
Social_Media	Yes	No	Yes
Property_Type	Apartment Building	Apartment Building	Apartment Complex
Days_to_Fund	2	0	1

For commercial properties we can observe that the higher target raise correlates more highly with failure when considering multi-family residential than it does for single family residential properties. And multiple buildings may also be a negative influence on funding success as the apartment buildings received funding in days vs. the apartment complex which failed.

4.7 Discussion

The results of this research indicate that backer's in a crowdfunding context evaluate real estate differently from the traditional financing method. Prior research in evaluating real estate in the traditional context placed a significant role of owner factors in the probability of a sale of a commercial building (Fisher et al., 2004). In a crowdfunding context, for residential properties, the borrower factors were not relevant. And for commercial, only the attribute of whether or not the borrower had social media accounts showed some correlation, but whether having these accounts was a positive or negative is inconclusive. This indicates that the unsophisticated "crowd" puts little importance on the borrower which suggests either the borrower is not an important risk factor or the fact that the platform has done due diligence is enough for the potential investor.

Prior research on evaluating real estate in the traditional context also placed a significant role of property factors in the probability of a sale of a commercial building (Fisher et al., 2004). Specifically, it finds that an increase in a property's age increases the sale probability and an increase in the property's square footage decreases the sales probability (Fisher et al., 2004). Older and smaller is better in the traditional context. However, our research indicates that property factors were the least relevant factors in a crowdfunding context. For both the residential and commercial properties, property factors like location, age, and size are the least relevant. This may indicate that the "crowd" does not have the experience to judge how property factors can affect the performance of the investment.

The results of this research also indicate that backer's for real estate crowdfunding has both similarities and differences from prior research in equity & reward crowdfunding. Prior research indicates that retaining equity and providing more detailed information about risks strongly impact the probability of funding success (Ahlers et al., 2015). Our research confirms that measure of risk, as indicated by LTV or ARV ratio, is relevant to residential backers, but not as important for commercial backers. Considering that the loan to value ratio is a somewhat commonly understood financial term and should be familiar to anyone who has applied for a residential home loan, it isn't surprising that it correlates as high as it does. In addition, perhaps after testing the signaling power of this number, POL prominently displays the ratio on the primary solicitation page of all properties.

Our research also confirms that measure of risk, as indicated by hold period, is relevant to both residential backers and commercial backers. In fact, other than target raise, it is the only attribute that correlates highly with both. This suggests that the duration the investment is held is quite important for the relatively inexperienced 'crowd."

Prior research across multiple platforms indicate the size of required funding is generally negatively correlated with success (Manonov et al., 2017). Our research confirms this as it is clear that target raise was the most correlated with the speed of funding.

Other research indicates financial factors are not the most important (Lukkarinen et al., 2016). Our research doesn't support this finding as financial factors are relevant for both residential and commercial backers but with some differences. APR is less relevant to residential backers and is more important to commercial backers. This is a little surprising because APR is another headline number displayed by POL and should be familiar to anyone who has a credit card. It reflects the ultimate return on the investment and yet isn't as important to residential investors as it is to commercial investors.

Our research does indicate that financial factors are less important to commercial backers who put some weight on social criteria as evidence by the correlation with social media. This indicates the "crowd" who invests in commercial properties is taking the borrower more into consideration.

4.8 Summary

Starting with data from Patch of Land, a leading real estate crowdfunding platform in the United States, we conduct extensive preprocessing of the data which results in a dataset with 16 attributes that is further split into two datasets, one for single family residential properties (190 records), and one for commercial properties (79 records). Each dataset undergoes feature selection by which the relative strength of each attributes′ effect on motivation is measured by a correlation technique between the attribute and the amount of days to fund the campaign. Next, the K-means clustering algorithm is applied to provide a partitioning and segmenting of similar properties together and the results were interpreted. Finally, we discuss the implications of the results.

Chapter 5. Conclusion

5.1 Research Summary

Crowdfunding is a rapidly growing method for funding real estate projects, and knowledge about the factors which motivate investors is needed within this context. Data mining has yielded valuable insights in prior crowdfunding research but there is a lack of research using data mining techniques for real estate crowdfunding. Findings are both similar and different from prior research on investor motivation and reveal a hierarchy of specific criteria that increase or decrease the probability and speed of funding success within a crowdfunding context.

5.2 Contributions

This research makes both scientific and practical contributions. First, it contributes to the theory of knowledge discovery for engineering, construction & real estate industry using data-mining techniques. Second, it contributes a method to provide decision support for owners, developers, managers and platforms on whether to use the crowdfunding model to raise funds and how to design projects and campaigns for funding success.

5.3 Limitations and Further Research

The main limitation is that the investor's motivations are not directly studied except through their behavior (participating in funding or not). The investor may be influenced by other variables that are not part of the platform data such as social network connections to the borrower, geographic proximity to property, pictures of the property, as well as information in attachments (such as appraisals) on the platform and links to data that are external to the platform. Investor surveys and other qualitative research methods could be part of further studies.

This research only considered the solicitation phase of real estate crowdfunding and the success for the borrower. Therefore, further study can research the success for the investor by looking at default rates of the debt. With data on defaults, data-mining can determine the relevant factors that influence a business' failure to meet their commitment to the investor.

In addition, because this research only looked at platform data, further study can use data-mining to look at macro-economic factors that may also influence the motivation of investors.

References

Ahlers, G. K., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. Entrepreneurship Theory and Practice, 39(4), 955-980.

Ahmad, P. H., & Dang, S. (2015). Performance Evaluation of Clustering Algorithm Using Different Datasets.

Apgar IV, M. (1997). U.S. Patent No. 5,680,305. Washington, DC: U.S. Patent and Trademark Office.

Baum, J. A., & Silverman, B. S. (2004). Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. Journal of business venturing, 19(3), 411-436.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. Journal of business venturing, 29(5), 585-609.

Cohen, J. (2016). A Study on the History and Functionality of Real Estate Crowdfunding. Joseph Wharton Research Scholars, University of Pennsylvania, PA.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. Journal of Management, 37(1), 39-67.

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

Fisher, J., Gatzlaff, D., Geltner, D., & Haurin, D. (2004). An analysis of the determinants of transaction frequency of institutional commercial real estate investment property. Real Estate Economics, 32(2), 239-264.

Gorunescu, F. (2011). Data Mining: Concepts, models and techniques (Vol. 12). Springer Science & Business Media.

Lukkarinen, A., Teich, J. E., Wallenius, H., & Wallenius, J. (2016). Success drivers of online equity crowdfunding campaigns. Decision Support Systems, 87, 26-38.

Mamonov, S., Malaga, R., & Rosenblum, J. (2017). An exploratory analysis of Title II equity crowdfunding success. Venture Capital, 19(3), 239-256.

Massolution. "2015 Crowdfunding Industry Report", <"www.crowdsourcing.org> (2017).

Misra, A., Gooze, A., Watkins, K., Asad, M., & Le Dantec, C. (2014). Crowdsourcing and its application to transportation data collection and management. Transportation Research Record: Journal of the Transportation Research Board, (2414), 1-8.

Rakesh, V., Choo, J., & Reddy, C. K. (2015, April). Project Recommendation Using Heterogeneous Traits in Crowdfunding. In ICWSM (pp. 337-346).

Silwattananusarn, T., & Tuamsuk, K. (2012). Data mining and its applications for knowledge management: a literature review from 2007 to 2012. arXiv preprint arXiv:1210.2872.

Tomczak, A., & Brem, A. (2013). A conceptualized investment model of crowdfunding. Venture Capital, 15(4), 335-359.

Valančienė, L., & Jegelevičiūtė, S. (2014). Crowdfunding for creating value: stakeholder approach. Procedia-Social and Behavioral Sciences, 156, 599-604.

Vogel Jr, J. H., & Moll, B. S. (2014). Crowdfunding for Real Estate. The Real Estate Finance Journal, 5-16.

Yuan, H., Lau, R. Y., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. Decision Support Systems, 91, 67-76.

Appendix A – Attributes and Descriptions of Data before Preprocessing

Attribute	Detailed Description
CFIIN	Crowdnetic Issue Identification Number.
Offering_Name	Company’s Name.
Portal_Name	Technology platform/funding portal created to showcase and facilitate these investments.
Sector	Area of the economy in which businesses share common characteristics.
Subsector	Subdivision of a given sector.
Industry	Subdivision of a given subsector.
City	City where the Company's headquarters or principal place of business is located.
State	State where the Company's headquarters or principal place of business is located.
Country	Country where the Company's headquarters or principal place of business is located.
Offering_Website	Company's website address.
Portal_Link	Intermediary's website link on the specific Company information.
Security_Type	Type of security – e.g. Convertible Debt, Debt, Equity, Real Estate, Revenue Sharing/Royalties.
Status	Active and Closed status available via market data service.
Reported_Start_Date	Start date of the offering Reported by Company.
Date_Added	Date the Company was added to CrowdWatch.
Women_Led_Offering	Any offering that has one or more C level or executive level woman in management.
Target_Raise	Total amount of the offering.
Total_Invested_Raised	Total amount of committed capital.
Min_Investment	Minimum Investment Amount.
Max_Investment	Maximum Investment Amount.
SM_URL1_Linkedin	Social Media LinkedIn page.
SM_URL2_Facebook	Social Media Facebook.

SM_URL3_Twitter	Social Media Twitter.
SM_URL4_GooglePlus	Social Media GooglePlus.
Purpose	A short description of the stated purpose and intended use of the proceeds of the offering sought by the issuer with respect to the target offering amount.
Property_Type	The type of property to which investments are directed, and may include, though not be limited to, industrial, commercial or residential.
Hold_Period	This is the anticipated amount of time that a property is held before exiting the investment.
Strategy	This is the intended strategy to add value to an investment.
Structure	This refers to the structure of the acquisition, such as debt, equity, loan to own, etc.
Projected_Cash_Return_Low	This is the lowest estimated return of an investment, represented as a multiple of the initial investment.
Projected_Cash_Return_High	This is the highest estimated return of an investment, represented as a multiple of the initial investment.

Appendix B – Details of Data Preprocessing

1. Data Reduction - Removal
CFIIN: Unnecessary for data-mining.
Portal_Name: Attribute removed because all from the same platform.
Sector: Attribute removed because all Financial.
Subsector: Removed because all Real Estate.
Industry: Removed because all Real Estate Development.
Country: Removed because all US.
Offering_Description: Removed because data captured in other fields such as Property_Type, Hold_Period and Strategy.
Borrower_Description: Removed because added an attribute showing years of experience because that is consistent info across records. Due diligence of POL only allows experienced borrowers, so differing factor is years of experience.
Offering_Website: Only three companies had a website that was not the intermediary website so was removed.
Portal_Link: Data from POL already represented in attributes.
Security_Type: Removed because all are loans. Some are purchase and some are refinance. This is more important distinction.
Status: Removed because all closed.
Date_Added: Crowdwatch is data collection company and irrelevant to the investor. The important date is start date of offering reported by company.
Women_Led_Offering: There was only 1 record where this was true, and it failed, which might distort the importance.

Max_Investment: Removed because all “None.”

Purpose: This information is in the project description.

Projected_Cash_Return_Low: Deleted because not shown on Patch of Land offering page.

Projected_Cash_Return_High: Deleted because represented in APR.

2. Data Reduction –Aggregation

If had 1 social media had all four, so reduced 4 attributes to 1 called "Social_Media" as a binary attribute: “yes” or “no.”

3. Transformation

Year_Built converted to Age-Years

Borrower_Description converted to Borrower_Experience_Years.

Days_to_fund blank values transformed to 0.

Min_Investment 5000 or 0 changed to “yes” or “no."

4. Remove Outliers

This was done using the weka preprocessing tool called interquartile range. The details in weka are “weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-last-O3.0-E6.0-w”

5. Converting from Numeric to Nominal

Numeric to Nominal is an unsupervised preprocessing tool from weka described as "filter for turning numeric attributes into nominal ones." This was applied to Days_to_Fund as a necessary step before can use correlation algorithms.

Appendix C – Feature Selection Run Information I

1. Single Family Residential Property Dataset (190 records)
=== Run information === Evaluator: weka.attributeSelection.CorrelationAttributeEval Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1 Relation: POL_Final_9_unbinned_remove outliers-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-last-O3.0-E6.0-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C17-Llast-weka.filters.unsupervised.attribute.Remove-R17-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C12-L2-V-weka.filters.unsupervised.attribute.Remove-R17-weka.filters.unsupervised.attribute.NumericToNominal-R16 Instances: 190 Attributes: 16 Evaluation mode: 10-fold cross-validation === Attribute selection 10 fold cross-validation (stratified), seed: 1 === average merit average rank attribute 0.098 +- 0.008 1.5 +- 0.5 15 Target_Raise 0.096 +- 0.004 1.7 +- 0.78 10 Hold_Period_Months 0.088 +- 0.004 3.1 +- 0.54 14 Strategy 0.082 +- 0.006 3.9 +- 0.94 9 LTV_or_ARV_Ratio 0.071 +- 0.004 5 +- 0.45 11 Min_Investment 0.06 +- 0.005 6.5 +- 0.92 8 APR 0.057 +- 0.003 7.4 +- 1.2 5 Borrower_Experience_Years 0.054 +- 0.001 8.8 +- 1.17 2 State 0.053 +- 0.006 9.1 +- 1.45 7 Borrower_Contribution_Ratio 0.052 +- 0.004 9.5 +- 1.75 13 Social_Media 0.049 +- 0.01 10.8 +- 2.18 6 Borrower_Contribution 0.047 +- 0.004 11.5 +- 1.5 4 Building_Age 0.044 +- 0.003 12.4 +- 1.11 3 Gross_Living _Area_Sq_Ft 0.039 +- 0.001 13.8 +- 0.4 1 City 0 +- 0 15 +- 0 12 Property_Type

Appendix D – Feature Selection Run Information II

1. Commercial Property Dataset (79 records)
=== Run information === Evaluator: weka.attributeSelection.CorrelationAttributeEval Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1 Relation: POL_Final_9_unbinned_remove outliers-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-last-O3.0-E6.0-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C17-Llast-weka.filters.unsupervised.attribute.Remove-R17-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C12-L2,7-weka.filters.unsupervised.attribute.Remove-R17-weka.filters.unsupervised.attribute.NumericToNominal-R16 Instances: 79 Attributes: 16 Evaluation mode: 10-fold cross-validation === Attribute selection 10 fold cross-validation (stratified), seed: 1 === average merit average rank attribute 0.129 +- 0.005 1.1 +- 0.3 15 Target_Raise 0.121 +- 0.007 2.4 +- 0.92 8 APR 0.113 +- 0.004 3.3 +- 0.78 10 Hold_Period_Months 0.112 +- 0.006 4 +- 1.73 13 Social_Media 0.107 +- 0.004 5.3 +- 0.78 12 Property_Type 0.102 +- 0.007 6.6 +- 1.56 5 Borrower_Experience_Years 0.101 +- 0.006 6.6 +- 1.69 6 Borrower_Contribution 0.096 +- 0.003 8.4 +- 1.02 14 Strategy 0.093 +- 0.009 9.6 +- 1.96 11 Min_Investment 0.093 +- 0.006 9.8 +- 1.72 7 Borrower_Contribution_Ratio 0.09 +- 0.009 10.4 +- 2.24 9 LTV_or_ARV_Ratio 0.088 +- 0.003 11.3 +- 1.1 2 State 0.079 +- 0.008 12.4 +- 1.56 4 Building_Age 0.064 +- 0.002 14.3 +- 0.46 1 City 0.062 +- 0.006 14.5 +- 0.81 3 Gross_Living _Area_Sq_Ft