Connecting the dots, an adaptable startup platform

Cristhian Bruno Ayllon Calderon
University of Lausanne
cristhianbruno.aylloncalderon@unil.ch

Valeria Medinaceli Molina
University of Lausanne
valeria.medinacelimolina@unil.ch

Martin Ruilova Quezada
University of Lausanne
martin.ruilovaquezada@unil.ch

***Abstract*—The Bolivian startup ecosystem represents a high-growth investment opportunity. However, the lack of adequate data and networks hinders the entrance of investors, an aspect that slowdowns the capital inflow. The development of an investors' platform that provides the key features of the startup market is paramount for its progress, aspect that is this research's focus. Currently, a comprehensive and complete database of the Bolivian startup industry does not exist; therefore, this research uses a database with comparable characteristics to develop the platform's structure. This platform's adaptability and replicability are essential to implement this solution once the Bolivian database is available.**

***Index Terms*—Bolivia, startup ecosystem, database, key features, platform, structure, adaptability, replicability.**

I. Introduction

Like many other developing countries, Bolivia has a flourishing startup industry with attractive potential to fulfill not only the investors' financial return requirements but also the social and environmental returns. The Bolivian startup's ecosystem, as in many countries, is characterized by highly motivated entrepreneurs, technology-driven ideas, and a steep learning curve to success. In most cases, Bolivian startups have a heuristic approach, belong to the informal sector, and are self-financed.

Many factors support the country's potential to become an attractive hub for startup development [1]. Software technology exports are strong, based on highly-skilled and creative software developers and low labor costs. Cochabamba had become the leading software producer and exporter in the country. This city concentrates several software-development companies, some of them working with global companies like Microsoft and training developers to provide offshore solutions for global businesses. E-commerce, e-banking, and mobile payments present a significant opportunity for Bolivian businesses to attract the continuously increasing percentage of the population with a mobile device with an Internet connection. One first attempt, on a national basis, was the 2018 second Christmas bonus, which was partially paid to each worker's mobile wallet forcing a large portion of the population to buy using a mobile application for the first time in their lives in many cases.

Innovation and technology represent vital opportunities for the country's development. Nations that understand the relevance of innovation in their economic agenda allocate a higher GDP percentage to "Science and Technology" and to their development agencies, which is a pending issue in the Bolivian case.

Therefore, despite efforts and advances made to make the startup environment flourish, the reality is that there are clear opportunities to enhance the Bolivian startup's ecosystem.

II. Research question

A. Problem

A current problem that the environment faces is the lack of information about the evolution of its main components, the startups. The last and only census made was in 2019 [2], which only reached 152 startups related to the technology sector, condensed on the most economically dynamic cities. The census was able to identify only three startup incubators and four accelerators in the country, all of recent creation. A crucial detail is the absence of any formal investors on its equity.

Besides, there are only a few research papers that were conducted on Bolivian startups. Most of them discuss too precise topics, for instance, the analysis made on the entrepreneurial ecosystem in the city of La Paz [3], which analyzes the principal drivers of entrepreneurs on a specific period. Hence, because the Bolivian academia did not deliver usable insights yet, startup cofounders are forced to rely on foreign literature or, in the simplest, adopt a conventional approach.

The private sector in Bolivia is, mostly, conformed by informal businesses. According to the data of Fundempresa, the commercial license provider, approximately 79% of authorizations of 2019 belong to "unipersonal" [4], the standard category among informal firms. The drivers of the latter situation are higher taxation and relentless workforce legislation. So, entrepreneurs prefer to stay informal until take-off and reach maturity, particularly considering that it takes at least 50 days to formally set up a new company due to the overly bureaucratic system. The difficulty of establishing new ventures puts the country in the last places in the World Bank's Doing Business Ranking [5], Bolivia classified as 150 among 190 economies in 2020.

Informality makes startups not subject to formal financing. Currently, there is no local venture capital (VC) or business angel networks in the country. The few local investors remain

very risk-averse and traditionally oriented, which is not in line with the highly dynamic, and therefore risky, startup environment. That is why angel investor's programs are highly demanded; otherwise, startup's equity would be self-financed, which in most cases is not enough to develop and to validate an MVP (Minimum Viable Product). The lawmaker, aware of the last, began developing a law to promote the sector. Nonetheless, the regulation can be solidified by a profound understanding of the ecosystem's fundamentals. Currently, Bolivia is the only Latin American country with no incubators or accelerators supported by the central government.

Of the several shortcomings mentioned, we find the lack of a proper connection between startups and local and foreign potential investors as the paramount next step to reinforce, where the other factors can rely on, to improve and strengthen the Bolivian startup ecosystem.

B. Objective

Our main goal is to provide both local and foreign investors with a comprehensive platform of the startup environment to address the main shortcoming found and help the ecosystem's growth by connecting investors and companies. Furthermore, we aim to be one step ahead by showing the essential features when analyzing a potential investment and, using those features, provide a prediction of the potential success or failure of a startup through our in-house developed model. This way, we aim to increase the interest of formerly reluctant local investors and specialized foreign investors in entering into this riskier yet attractive sector.

However, the development of an investors' platform for the Bolivian startup market requires the construction of a comprehensive database of the environment. As we observed in our preliminary assessment, the market lacks a public comprehensive and complete database of the existent startups in the country. The construction of this database represents a significant effort and an extended period, due to it requires the participation of many of the current actors. This process requires contacting those startups already included in the first census, reaching the companies not discovered in that attempt, and find new companies that had been created in the meantime. This task has not started yet. Nevertheless, the creation of the platform's structure and functionality is a parallel process that can be carried out independently of the creation of the Bolivian startup database.

Besides, our society is facing times of uncertainty that force the different market agents to provide out-of-the-box solutions for its growth. So, the adaptability and replicability of these solutions is a paramount feature to generate a real change in this subject.

Therefore, this project's objective is to develop the structure of the investors' platform that displays the most relevant market information as well as our prediction model. This process uses a publicly available database as an example, resulting in a template structure as the first outcome. This structure is to be adjusted for its final implementation in a further stage once the Bolivian database is available.

C. Scope

Considering the information constraints that we are facing off, this project's scope lies strictly in the development of the platform's template structure based on the example database. This deliverable shall be able to provide the relevant market insights and perform the prediction algorithm according to the input database.

III. Methodology

The first step to pursue the project's objective is to find a startup database to serve as an example for the construction of the platform's template structure. We selected the CAX_startup database uploaded to Kaggle. The raw dataset consists of 472 entries, each representing a different startup with 116 features for each one. We choose this set as an example base because its characteristics are similar to what we could receive when the Bolivian database is constructed. These similitudes are its small number of entries, considering the existence of a limited number of startups, and a consistent set of features (more than 100), selected to provide a more in-depth insight of the market considering different approaches.

We have faced this issue on a two-part project consisting of a Prediction model and Investors' Platform.

A. Prediction Model

This section lies within the scope of another research project developed in parallel by our team. In summary, this project uses the same example database to structure a methodology able to select the features that have the most significant influence on the success of startups, given the available information on the dataset. Then, using these features construct a classification model that can predict the potential outcome of a startup, which could be a success or failure. It is important to remark that success in this context is defined as a startup that raised capital and is still in operations, and vice versa; failure states as a startup that went to bankruptcy or had to close its operations.

This way, we can provide the market with a more in-depth insight into the early stage of a startup and provide the investors with a specialized tool to ease the screening part of their due diligence process. Therefore, investors can focus their resources more intelligently, and the entrepreneurs can redirect their efforts to reinforce these features.

B. Investors' Platform

The present project is in charge of the development of this section. Through this process, we aim to develop the platform's structure that displays the industry's most relevant

facts and allows the investor to search for more detailed information by sub-industry to make a more informed investment decision. Furthermore, this structure includes the prediction model, developed in parallel research, to enable investors and entrepreneurs to interact with the model and improve their prospectives.

We applied Python as a coding language for the development of the structure, and Tkinter for the Graphic User Interface (GUI) construction. The implementation of the code required to create the platform is explained in the section below.

IV. Implementation

The project is hosted in a Github repository, which contains all the required files and instructions on a step by step basis. Please refer to the README for the adequate installation of the Kaggle API, environment, and packages required for the proper replication of the platform. The program runs in Python, version 3.8.2. and is structured using the following packages:

- Tkinter - version 8.6.8
- Numpy - version 1.18.1
- Pandas - version 1.0.3
- Matplotlib - version 3.1.3
- Scikit-Learn - version 0.23.0
- Pillow - version 7.1.2
- Pandastable - version 0.12.2

A. Pre-processing algorithm

Pre-processing the database is an essential step for the structuration of both the model and platform. Throughout this process, we have first analyzed the data and then proceed to uniformize each feature and to clean misspelled words or out-of-range values, always considering the variable's nature. All this work was implemented through a pre-processing algorithm written in the Jupyter Notebook named 'Preprocessing_PROG.ipynb'.

The example database is obtained directly from Kaggle using its API. The pre-processing algorithm first downloads the raw database in a CSV file. Then, the program uniformizes the empty entries and then deletes all non-useful features. The next step is to correct typo mistakes, capital missuses, space problems, redefine industry categories, and rename all the headers for a uniform presentation, to provide finally with a clean output called 'CAX_Preprocessed_PROG.csv'.

The processed database comprises 472 startups with 102 features, which is the primary input for the platform construction.

B. Program implementation

The platform code has a modular structure presented as follows (we only display the most important executable files for diagramming reasons):

```
pro_project
├── README.md
├── PRO_Final_Project_VMB.py
├── Preprocessing_PROG.ipynb
├── pro_project_env.yml
├── VMB_Generators/
│   ├── Button_Generator.py
│   ├── Graph_Generator.py
│   ├── Industry_Selection.py
│   ├── KeyFacts_Sector_Generator.py
│   ├── Model_Generator.py
│   └── Pivot_Table_Generator.py
└── .git
```

The main script is written in the file called 'PRO_Final_Project_VMB.py' placed on the repository's root directory. This script runs the primary structure of the platform that is explained below. To run this script appropriately, we rely on the following internal dependencies, hosted in the folder ṾMB_Generators/. The dependencies and their functions are:

1) 'Button_Generator.py': designed to generate all the buttons that connect every frame of the platform.
2) 'Graph_Generator.py': designed to create all the graphs.
3) 'Industry_Selection.py': designed to ease the selection of sub-industry and economic activity for the VMB Model.
4) 'KeyFacts_Sector_Generator.py': designed to display the key facts of each sector on the platform.
5) 'Model_Generator.py': designed run the prediction algorithm of the VMB Model.
6) 'Pivot_Table_Generator.py': designed to generate all the tables.

Each dependency hosts the required functions to structure the platform and the widgets needed for the GUI.

C. Structure

The user interface has the following structure, displayed graphically in the flow chart in Fig. 1:

1) Main page: Platforms main window.
2) Dashboard: Relevant information of the database displayed as graphs and tables.
3) Analysis per sector: Presents relevant features of the database on an economic sector basis.
4) VMB Model: Prediction model of the success of a startup based on the algorithm we generated through the Advanced Data Analytics project.

We have created a brand called 'VMB Venture Capital' for preliminary presentation purposes. Therefore, the platform, model, logo, and methodology take this pilot name within this project's scope. The name does not refer to any real company or association whatsoever. Any similarities to any real organization are purely coincidental.

1) Main page: Platform's main window. It displays the most relevant facts from the database and the options to move

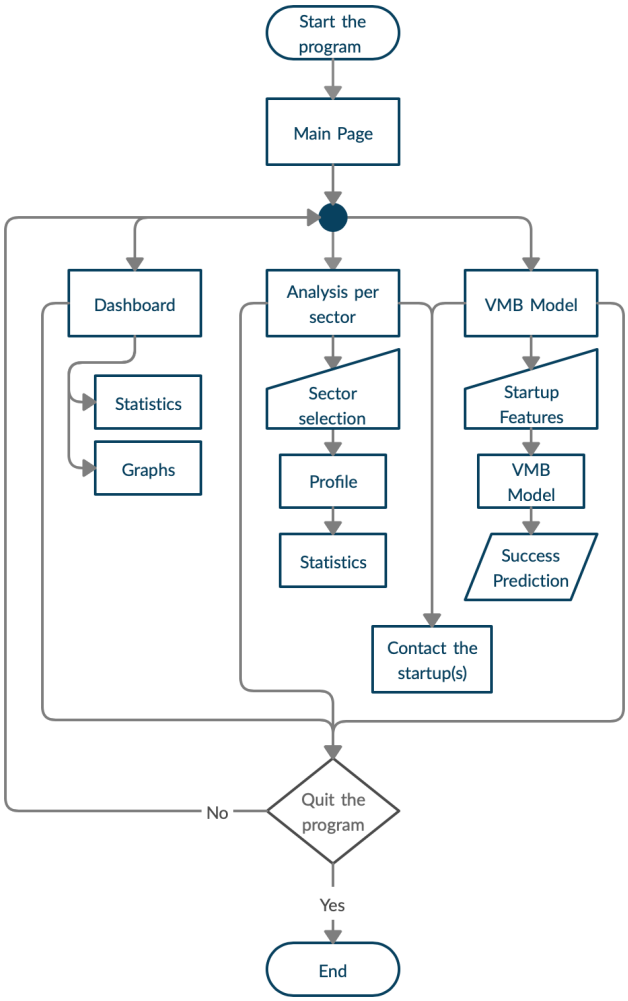

Fig. 1: VMB - Platform's Flow Chart

around the program. This section consists of one frame which shows six labels for the database's principal figures and the respective buttons to connect to each section on the platform.

2) Dashboard: This section's front page displays a plot pie with the percentage of successful and failed companies on the database and the buttons to connect to the two inner sections:

1) Statistics: Consists on a table classified on an economic sector basis and that displays each sector's number of companies, success rate (in percentage), average las funding amount received in USD millions, and the number of seed investors, angel or venture capital investors, and advisors.
2) Graphs: Consists of two frames, each with four different graphs presenting the most relevant factors that explain the database.

The facts displayed in both the table and charts rely on the key features found in the prediction model development. Thus, they intend to show the database distribution in terms of the characteristics that could lead to a startup to be or not successful.

3) Analysis per sector: This section's objective is to provide a better understanding of the different sectors of the economy where these startups work. We assessed the activities of each startup and classified them according to the Global Industry Classification Standard (GICS), which divides companies into eleven sectors, 25 industry groups, 71 industries, and 165 sub-industries.

Currently, the example database (CAX_startups) does not have companies working either in the Materials nor Consumer Staples sectors and is limited to 13 industry groups, 18 industries, and only 25 sub-industries; thus, the current structure is designed to reflect only this database composition. However, its design makes it easily adjustable to incorporate any other new economic activity and correctly classify it into its respective GICS categories, when implementing the structure for the Bolivian use.

The analysis per industry is structured as follows:

1) Frame 1: This section displays the startup ecosystem classified by economic sector. It is worth mentioning that the classification respects the GICS framework; therefore, it centers not only on the type of product/service to be delivered but also in the specificity of the clientele. The latter means that, for example, if a startup is a tech solution for the real estate market exclusively, it is to be classified within Real Estate Services.
2) Frame 2: This screen includes:
    a) A plot of the average amount of funding received by successful and failed startups in each sector. As the database is feed with more entries, this illustration can provide insight into the expected amount that could be raised on a sector basis.
    b) The option to select a specific sector to analyze.
3) Analysis per sector:
    a) Once the user has selected and submitted a sector, the platform displays the sector's profile. This profile consists of:
        i) A plot of the average amount of funding raised by the sector's startups, classified on a sub-industry basis and according to their success status. As mentioned earlier, this illustration can provide more precise insight as the database is feed with more recent entries. This way, it can become a benchmark for potential ticket sizes for the investments in each sub-industry, according to its specific characteristics, requirements, and risks.
        ii) Key facts of the sector, including the number of companies currently present, companies' location (countries in the case of the example database), the success rate of the startups in the

sector, the average amount of funding raised, and the average number of investors (seed, angel and venture capital) and advisors per startup.

iii) A button to raise the sector's statistics frame. This frame displays a table of the companies in each sector classified on a sub-industry basis with similar information as the main statistics table presented in the Dashboard.

iv) A button to contact the startups of the selected sector. Once the investor reaches a sector of her interest through our platform, she can choose to receive detailed information about the targeted ventures and their contact information. Due to the available information on the example database (CAX_startups), we cannot provide the contact information within the scope of this project. However, this platform's feature can be fully implemented once we can use the Bolivian database and considering all the confidentiality issues related to sharing a company's information to a third party. This factor is pending to be considered and developed on a further stage.

The platform's structure requires the user to submit its selection, through the button 'SUBMIT', every time she changes its selected sector; this way, the platform displays the adequate frames. If the selection is not appropriately submitted, the platform presents the information of the last selected sector or the Dashboard if there was no selection submitted previously.

4) VMB Model: The model research and development is carried out in a parallel process by our same team and using the same example database. The scope of such work is to provide a methodology to find the most relevant features and prediction model using a given dataset. Thus, during this research stage, these findings respond solely to the CAX_startup database used as an example in the pilot implementation.

The VMB Model section displays the feature selection options, and the prediction made using our classification model. This segment is structured in four frames, three for the feature selection and one to present the results.

Using the example database (CAX_startups), the methodology generated through our parallel research provides a set of 12 most relevant features for the prediction. We classified these features into four categories:

1) Industry. This section includes the following variables, which can be selected in the segment's first frame:
   a) GSE - Sector
   b) GSU - Sub-industry
   c) IND - Economic activity
2) Experience of the team. This category is part of the second frame and consists of the following variables:
   a) AYE - Average relevant experience of cofounders
   b) BIG - Cofounders have worked on one of the Big 5 consulting firms
   c) REC - Number of recognitions to cofounders
   d) SUR - Survival through recession periods of cofounders' previous startups
   e) INC - Investment received through global incubator competitions
3) Market Characteristics. This feature is presented in the third frame and includes the following variables:
   a) BAR - Existence of barriers of entry to the industry
   b) DWF - Difficulty of obtaining workforce
4) Business model. This category is located in the segment's third frame as well. It includes the following variables:
   a) PST - Existence of a pricing strategy
   b) CPB - The service or product based on a platform or cloud
   c) CDA - Focus on consumer data
   d) MLB - Machine learning base

Once the user selects the values in each feature, the platform leads to the VMB Results frame. This section initially presents one single option, 'Charge Inputs,' which presents all the features and the user's selections made in the previous frames. Once the user charges her entries, the platform automatically presents a 'Get Results' option, which runs the prediction model. This process displays the startup's potential outcome prediction, resulting in either 'Success' or 'Failed,' and the precision at which is calculated. Finally, this frame presents a brief explanation of the methodology, scope of both research projects as well as the 'Contact the startup' option, similar to the one presented on the analysis per sector.

It is critical to remember that when using the platform, it is required to push the 'SUBMIT' button every time the user decides to change one or many of the inputs selected. If the new selection is not submitted, the program uploads the last submitted information to the prediction model. The 'Charge Inputs' option allows the user to see the last submitted options if she desires to modify one of them before the prediction.

On a side note, we present an option called 'Features Explanations' located in the left-down corner of each frame within the VMB Model section. This option presents the user a PDF document (opened in her preferred browser) that provides more in-depth information on the model's features, the possible values that each variable can take, and its meanings.

It is worth mentioning that the features included in this process, that are used and explained throughout the model implementation respond solely to the outcomes received using our methodology with the example database. The features, explanations, and the model per se, are prone to change

when using a different database, as both projects are meant to provide the methodology and platform's structure and not the exact results of the given database (CAX_startups).

V. Maintenance and Update

The outcome of this project's code implementation is the platform's structure, which is based entirely on the example database in this pilot stage. Therefore, the proper maintenance and update of this code are paramount for the next step of the development of our proposed solution of the shortcomings identified.

The codebase for both parts of the global project (investors' platform and prediction model) is updated in their respective Github public repositories, which allows the researchers to keep the projects' code stored adequately for further advances and share these findings with other interested developers. All the code is appropriately commented and documented to ease its use by an outside party.

The adaptability and replicability of this work are essential to accomplish the global objective. Therefore, the project's code must be updated to adjust the structure to the incoming information, once the database for the Bolivian startup ecosystem is available. And then, a next step may be the adaptation of the platform to the investors' interests in terms of usability; that means, consider a potential implementation of the platform as a website or mobile application, and the development of a bilingual structure in Spanish and English to adapt the platform for its use by local and foreign investors (currently the platform is in English for research presentation purposes). All these potential implementations shall be considered on a further stage.

VI. Results

The implementation of this project's scrip generates the Graphic User Interface (GUI) of the investors' platform. The platform welcomes the user on the main page.

A. Main page

The main page of this structure is presented in Fig. 2. The development team designed the platform intending to be interactive and intuitive to ease the investors' use. Therefore, from the main page, the user can choose to go directly to each of the three sections of the platform Dashboard, Analysis per sector, and VMB Model, and also, the user can jump around the different sectors while using them.

B. Dashboard

The Dashboard welcomes the user with the plot of the successful and failed startups on the database, as observed in Fig. 3.

From this front page, the user can direct to the Graphs and Statistics sections, the latter divided into two frames. These screens are presented in Fig. 4 and Fig. 5.

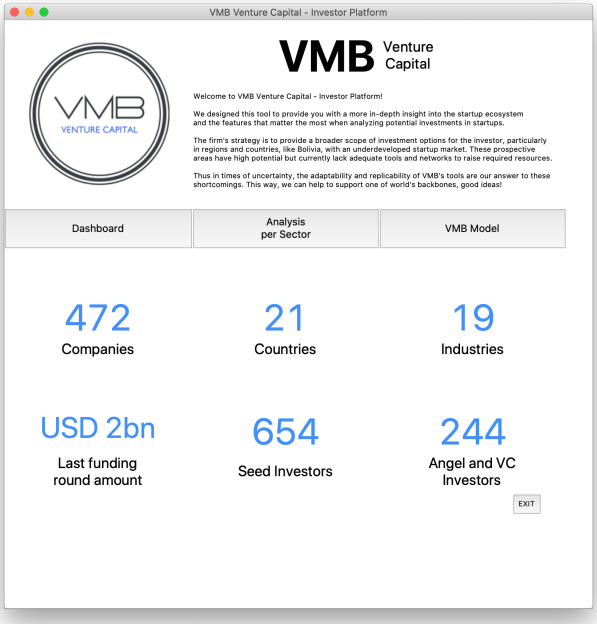

Fig. 2: VMB - Main Page

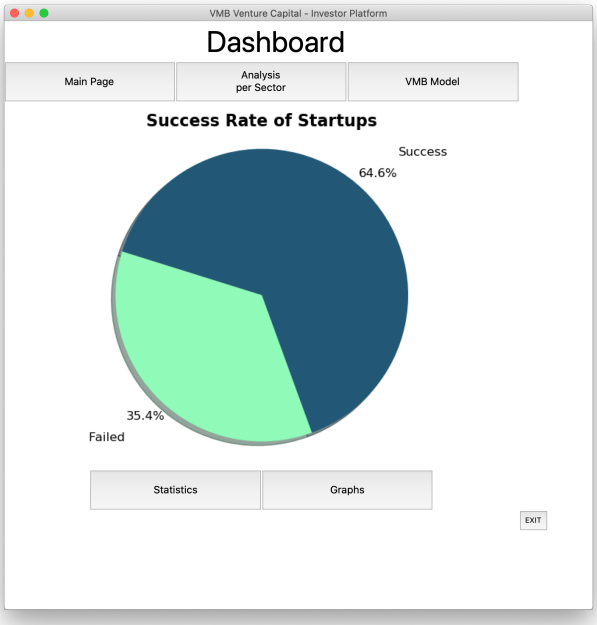

Fig. 3: VMB - Dashboard

C. Analysis per sector

The second segment of the platform presents two principal frames as per can be observed in Fig. 6. The section's second frame allows the user to select the economic sector she wants to analyze further. Once she selects and submits its choice, the platform directs the user to the sector's profile. Currently,

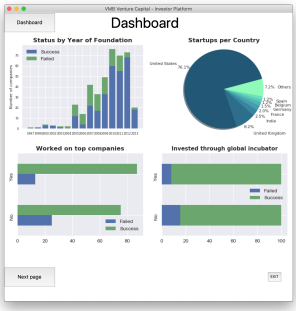

Fig. 4: VMB - Dashboard Graphs

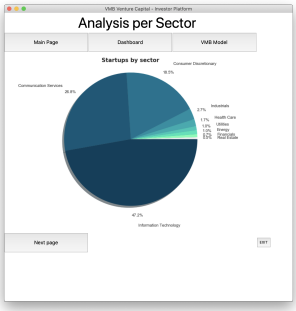

Fig. 6: VMB - Analysis per sector

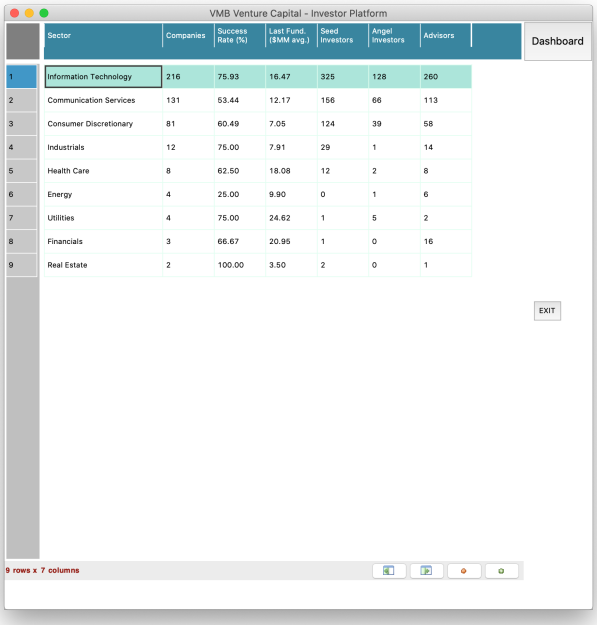

Fig. 5: VMB - Dashboard Statistics

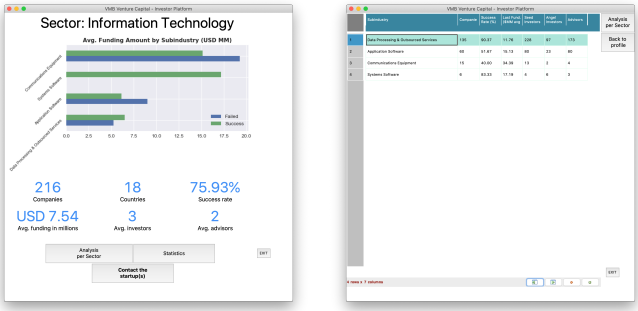

Fig. 7: VMB - Information Technology Sector

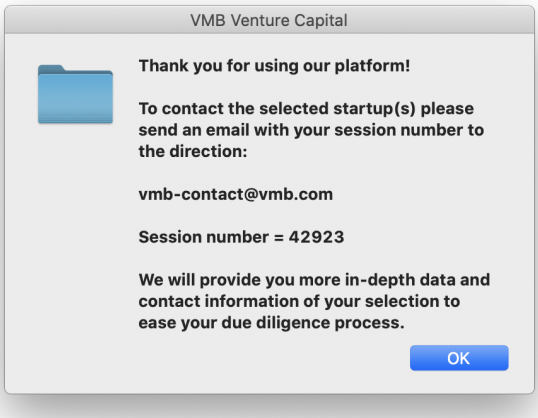

Fig. 8: VMB - Contact Startup(s)

the pilot version of the platform only displays nine sectors according to the GICS classification due to the example database participants.

When in the profile, the user can go back to the selection frame, to the statistics table, or contact the startup(s) of the selected sector. A preview of the sector's profile and statistics table is presented in Fig. 7, using the information technology sector as an example.

If the user requires to contact the startup(s), the system provides a message box with further instructions to follow; this message is presented in Fig. 8.

D. VMB Model

Finally, the last section of the platform allows the user to introduce the startup's features to run the prediction model. As explained in the implementation process, this section is divided into three frames for the value selections and one for the results. The feature selection frames are displayed in Fig. 9 and Fig. 10.

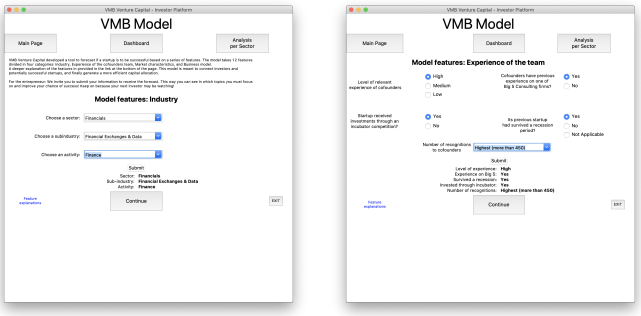

Fig. 9: VMB - Features Frame 1 and 2

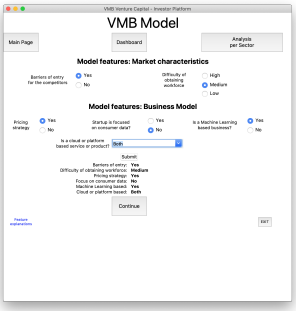

Fig. 10: VMB - Features Frame 3

The last frame of the platform (Fig. 11) presents the prediction results as well as a brief explanation of the model. This frame also allows the user to receive the contact information of the startups, as presented in the previous section.

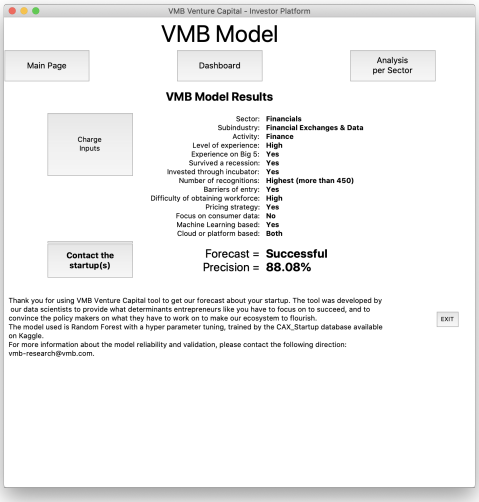

Fig. 11: VMB - Model Results

VII. Conclusions

There exists an attractive growth potential in the Bolivian startup ecosystem, a potential that could lead not only to the generation of value for the prospective investors, but generate employment, innovation, and boost the research and development in a country that requires sustainable ways to strengthen its economy and the improve the people's life quality. Therefore, solutions to improve the entrepreneurial environment are on high-demand in the Andean country.

We have identified that one of the many shortcomings is the lack of funding sources to boost the venture's development in the early stages. Local investors remain risk-averse and traditionally focused, while specialized foreign investors have not yet landed in the country. One of the main drivers of this problem is the lack of an adequate source that displays the ecosystem's relevant figures and properly connects investors and startups.

Therefore, this research is aimed to address this shortcoming through the development of an investor's platform to provide relevant insights that should be considered while analyzing the industry; this tool seeks to ease the selection part of the investment process.

Currently, there is no comprehensive and complete database of the Bolivian startup market, and its construction represents a long period and a significant effort of the many parties involved. However, we considered that the elaboration of the platform's structure is a process that can be made in parallel, to have this product ready for update and implementation when the national database is available. Missing information represents an obstacle, but there are options to consider in parallel, as the one presented in this research.

We selected an example database (CAX_startup from Kaggle) for the platform's pilot implementation. This database is the one that resembles the most with what we expect to receive from the Bolivian database construction. Using this example database, we created the platform's structure, and, in simultaneous research, we also developed a methodology to select which are the most relevant features in determining the potential success or not of a startup. Then, we developed a prediction model of this potential success using these features and machine learning algorithms. The inclusion of these features and the prediction model in the platform represent our added value for the investor.

The platform's structure is divided into three parts: Dashboard, that displays the entire database's most relevant features; Analysis per sector, a section that presents the relevant figures on an economic sector basis; and VMB Model, a segment that allows the investor (or maybe a concerned entrepreneur) to submit the characteristics of a startup within our set of selected features, and perform the

prediction of the potential success of the startup through our in-house created model.

The scope of this research is limited to the development of the platform's structure, considering the information constraints we are facing. In a further stage, it is essential to work, alongside with other interested parties, on the construction of a comprehensive and complete database of the startup environment. This factor is critical for the adaptation of our platform using real Bolivian data, furthermore, having a publicly available dataset could also encourage other participants in the industry to take action and be part of our market's potential success.

REFERENCES

[1] Article: "10 Facts That Showcase Bolivia's Technology Potential" https://www.launchwaymedia.com/blog/facts-that-showcase-bolivias-technology-potential

[2] Report: "Bolivian Startup Environment Baseline 2019" https://mapeoticbolivia.org/homepage

[3] Research paper: "Ecosistema del Emprendedor Paceño: Un Análisis Parcial (Volumen 3)" https://bit.ly/3g5oJ26

[4] Website: Fundempresa - Statistics 2019 https://bit.ly/2LOKXaP

[5] Report: "World Bank - Doing Business Report 2020" http://documents.worldbank.org/curated/en/688761571934946384/pdf/Doing-Business-2020-Comparing-Business-Regulation-in-190-Economies.pdf

[6] Article: "How to create a favorable environment for startups in Bolivia" https://lexlatin.com/opinion/como-crear-un-ecosistema-favorable-para-las-startups-en-bolivia

[7] Article: "The rising Bolivian startup ecosystem" https://latamlist.com/the-rising-bolivian-startup-ecosystem/

[8] Article: "Bolivian Startup Ecosystem" https://www.nathanlustig.com/bolivia-startup-ecosystem/

[9] Article: "Bolivian startup ecosystem innovation seeds are planted but waiting for water" https://medium.com/seedstars/bolivian-startup-ecosystem-innovation-seeds-are-planted-but-waiting-for-water-367bed863cb8