Where does the money in science go in Israel?

Lin Peretz

Israel science foundation (ISF) Grants provide some of the main funding sources for the field of science. This organization's website contains a wealth of information about the grants over the years, including details about the number of projects, the funds given, the fields, and the different uses. Analyzing the above information using quantitative methods, visual representations, and various text analyses is the goal of this project .The analyzes we conducted allowed us to obtain information about the distribution of the funds, the distribution of the various uses, the institutions, the central researchers, and the nature of the winning grants.

ISF | Text analyses | Grant

Interdiction

Israel is known for its robust science and technology sector, with a significant amount of funding being invested in research and development. According to data from the Central Bureau of Statistics, Israel's gross domestic expenditure on research and development reached 52.7 billion NIS (around 16 billion USD) in 2020, representing 4.9 percentage of the country's gross domestic product (GDP) 1.

The Israeli government is one of the primary sources of funding for science and technology research, with the majority of funding being provided by the Ministry of Science and Technology, the Ministry of Defense, and the Israel Innovation Authority. In addition, there are several private foundations and philanthropic organizations that fund research and development in Israel.3

One of the most significant sources of funding for science and technology research in Israel is the Israel science foundation (ISF) Grant program , is a key initiative of the Israeli government to support innovative research and development in Israel's science and technology sector.

This program is designed to promote economic growth and competitiveness in Israel by supporting the development of innovative technologies in a range of fields, including biotechnology, nanotechnology, cybersecurity, and renewable energy. The program offers several different types of grants, including research and development grants, pilot grants, commercialization grants, accelerator grants, and collaboration grants and has a budget of approximately 650 million NIS (around 200 million USD) per year and has funded more than 2,000 projects .2

Prior works were research Grant money, various uses, various effects and more, Meirmans S provides a systematic review of the literature on the allocation of grant money in scientific research, examines the factors that influence the allocation of grant money, the distribution of grant money across scientific fields, and the effectiveness of grant money in promoting scientific progress.6 Another study examines analyze which firms receive Research and development project grants in Germany and how this public support evolves over time by considering in particular firm's previous participation7.

Using data from Swiss National Foundation , another study examines the impact of a grant program promoting international mobility on researchers' scientific outcomes and career.8

As a result of the data files, we were able to extract a lot of information, including the distribution of grant money data among different fields (social science, life science, exact science, etc.), the number of winning projects in each field, the distribution of the money over time and even the distribution of the money to institutions and researchers, answering the question of where money goes in science in Israel, and performing additional analysis that includes the use of text analysis tools, such as Bertopic 5 and create Prediction model in order to predict the grant filed from the grant title.

In addition, this information was combined with information from Science-wide author databases of standardized citation indicators, publicly available database of over 100,000 top-scientists that provides standardized information on citations, h-index, co-authorship adjusted hm-index, citations to papers in different authorship positions and a composite indicator.4 As a result of this data connection, an analysis of the institutions that receive the most money has been added.

Method

Data. ISF website provides a wealth of information related to the ISF Grant program, including data on the program's budget, funding priorities, application process, and success rates, the website serves as a valuable resource for researchers, entrepreneurs, and companies interested in applying for funding through the ISF Grant program.

In this project, we used two central databases that the site provides, excel files that contain a lot of information about the distribution of funds and pdf files that also contain a discussion and an annual account for the years 2012-2022. Using this data allows us to perform a broad analysis of the information and get a situational picture of the distribution of the lanes, different institutions, areas, etc.

PDF Analyzed. ISF provides PDF files containing various text data about the winning grants over the years. All the PDF files of the years 2012-2022 contain a table that summarizes the distribution of allocations for grants in the core programs. Furthermore, these tables were in the exact same format (Fig 1), so we were able to extract data from them using a package that extracts text from PDF files.Based on this data, we can get a general picture of how funds were distributed across the fields over the years.

חלוקת ההקצבות למענקים בתכניות הליבה לשנת תשע"ג 2012/13 (באלפי ₪)

	חדשים		ממשיכים	
	מספר מענקים	סכום מענקים	מספר מענקים	סכום מענקים
מענקי מחקר אישיים				
מדעים מדויקים וטכנולוגיה	161	32,132	431	84,132
מדעי החיים והרפואה	167	37,250	390	88,433
מדעי הרוח	71	6,982	117	12,826
מדעי החברה	99	12,248	144	18,271
סה"כ	**498**	**88,612**	**1,082**	**203,662**

Fig. 1. Example of a table from a PDF 2012/13 file.

Excel Analyzed. There are five Excel files available on the ISF website which provide information about the grants in different categories (equipment, publications, post-doctoral, research, research centers, workshops). Each file contains information about each grant project, such as the core field, researcher's name, research title, total money, institution, and so on, by analyzing all of these, we could identify the distribution of broad information for example about the distribution of the money to institutions, core fields and researchers.

Due to the presence of the institution's name in this data, we were able to combine it with data from Elsevier 4, which provides additional details about researchers and their institutions ranking. Besides the quantitative analysis, we also performed a text analysis to obtain additional information about the various Grant projects, These analyses were performed using the research title field in the various files, which we translated into English so we could use the existing tools more conveniently.

We performed various analyzes such as using WordCloud 9 and Bert topic, we use TfidfVectorizer 11 to create a RandomForestClassifier 10 model that, given the grant title, can determine which of the four core areas (Exact sciences and technology, Life sciences and medicine, Humanities and Social Sciences) the grant belongs to.

Results

PDF Results. The data we extracted from the PDF files were used to analyze the distribution of budgets throughout the years. Our visual display shows how much money was given to each core area over the years, along with how many projects were funded in each core area. (Fig 3)

Over the years, most of the funds have been directed towards life sciences, medicine, and exact sciences and technology, accordingly, these fields receive the most grants out of the four fields.

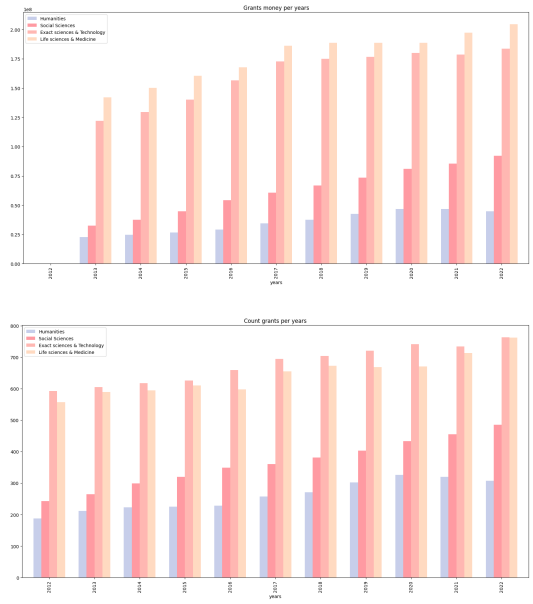

Fig. 3. Distribution of money / amount in the various fields over the years 2012-2022

Excel Results. Several types of analyses were conducted on the Excel data:

1. **Quantitative analysis**: The data were visually presented in several different ways. The first was the distribution of how money was used in all the different fields, for equipment, publications, post-doctoral, research, research centers and workshops (Fig 4. According to this analysis, most of the funds are intended for research centers in all fields.

   In the second section, we presented the distribution of money and the amount of grants awarded to the top 10 institutions (Fig 6). As can be seen from this analysis, the “Hebrew University in Jerusalem” is the leading institution in terms of quantity and money.

   And finally, similar to the previous analysis, we performed the same quantitative analysis of money and amount of grants awarded to the top 10 researchers as well (Fig 8).

2. **Rank Elsevier**: To examine the ranking of the top 10 institutions, we used the information we derived from Elsevier to present their ranking visually (Fig 9) . The highest ranking in this analysis goes to the “Technion”.

3. **Title analysis:** We chose to perform text analysis in several different ways in this part of the project, which allows us to gain additional information about the nature of the winning grants by analyzing their titles. To make existing text analysis tools more convenient and easier to use, all titles have been translated into English.

   (a) **Word Cloud:** In order to gain a better insight into the common words in the titles of the winning grants, we used a word cloud. (Fig 10) This was done after cleaning the stop words in all the titles.

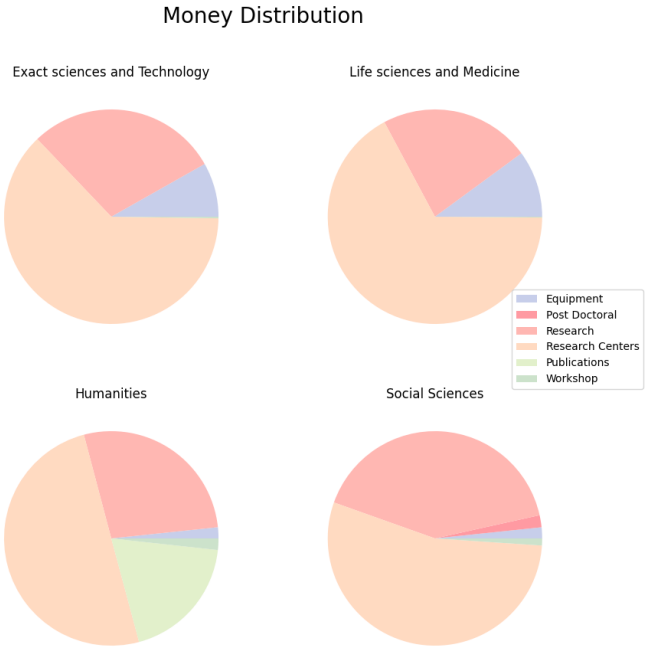

Fig. 4. The distribution of the use of money in the various fields

Fig. 10. Word cloud result for the grant titles

(b) **Bert Topic:** Secondly, we chose to use Bert topics for Topic Detection with class-based TF-IDF procedures. We were able to visually extract the 10 main themes in the winning grants using this analysis.(Fig (11)

(c) **Prediction Model:** Our final goal was to develop a classification model that would receive the grant title and be able to predict the core field in which the grant belongs. In this way, it is possible to determine how much the title indicates the nature of the various core areas. A fairly good classification model, with an accuracy of 0.52, was constructed using the TD-IDF method, a Random Forest classification model, and hyper-parameters. The confusion matrix of this model shows that, among other things, it predicts the best in ”Humanities” but also prefers it in many cases where it is not the right field. (Due to the large number of grants awarded in this field, we can assume this will happen.) Moreover, it is evident that the model cannot accurately identify the titles of the social sciences.(Fig 12)

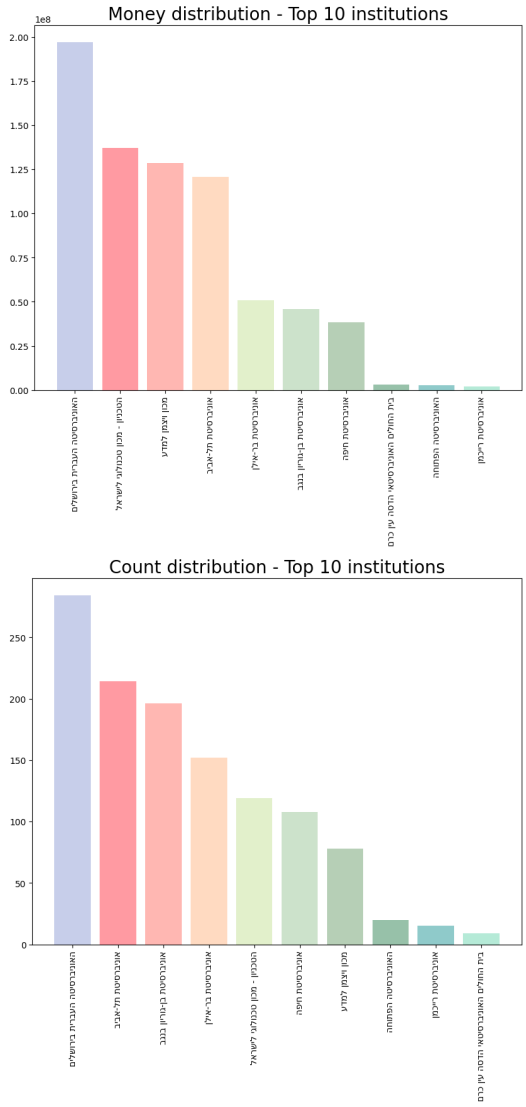

Fig. 6. Distribution of money / amount of the Top 10 institutions

Discussion

In this study, we analysed the data and we were able to draw many conclusions regarding Israel's grant programs.

To begin with, we can say that most of the funding for science in Israel has been allocated to the field of life sciences and medicine over the years, which suggests that in most cases the amounts have been increasing, and the distribution of funds between the various fields has remained very similar.

Furthermore, in all fields, most funds are allocated to research centers, and the largest amount belongs to the Hebrew University in Jerusalem. (which is ranked 3rd according to the analysis of the ratings we did). As compared to first place in the rankings, the "Technion", which also receives a considerable amount of money (second place), but not necessarily for a high number of projects, the amount per project is probably higher.

Based on our analysis of the researchers, Professor "Haim Sider" received the largest amount. From searching his name on the Internet, it appears he belongs to the Hebrew University, which is in line with the previous analysis.

The analysis of the research titles shows that there is a large number of topics (4 out of the top 10) that relate to the world of medicine and life sciences (diseases, cancer, bacteria, pain, trauma, etc.). According to the analysis, most of the money is directed to these fields.

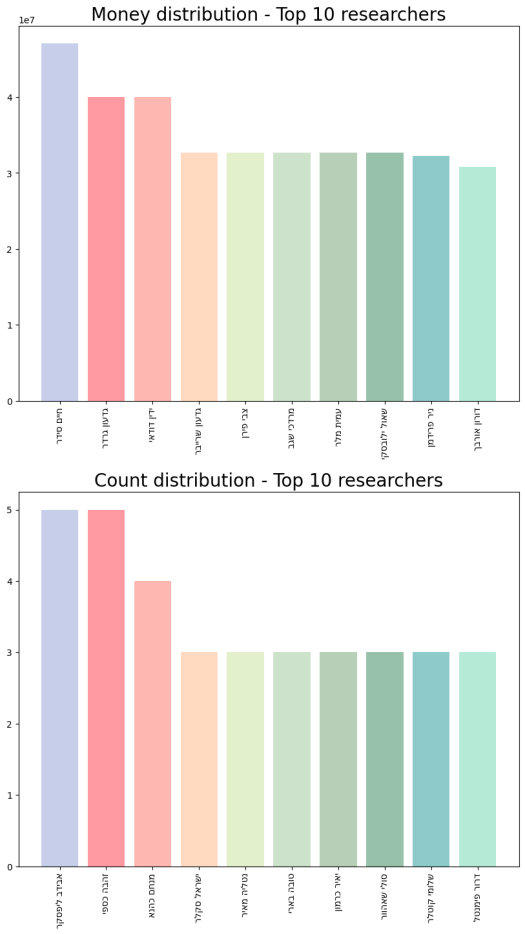

Fig. 8. Distribution of money / amount of the Top 10 researchers

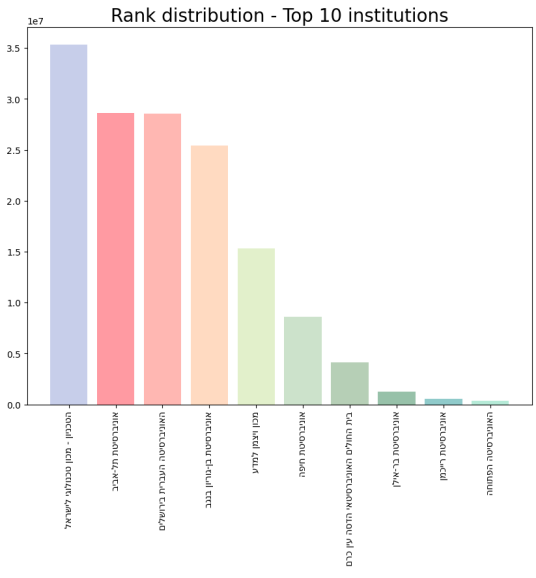

Fig. 9. The ranking distribution of the top 10 institutions

In conclusion,the ability of the classification model to be broken down into core areas by title indicates that it can be used to learn and understand how the areas work.

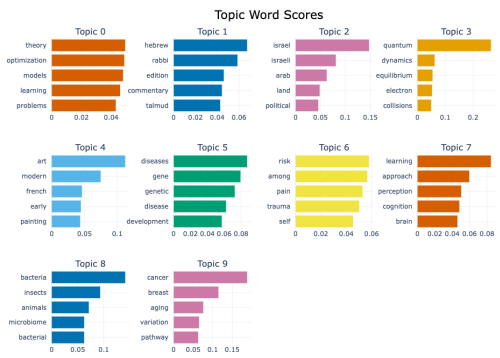

Fig. 11. Bert Topic Top 10 Word Scores

Fig. 12. Confusions matrix of Random Forest classifier model of predicting the core field from the title

Conclusion

In this project, quantitative analysis, ratings analysis and textual analysis were performed on the winning Grant data in Israel. This analysis provides a very good overview of where the money in science goes in Israel, the distribution of the money over the years, the common uses of the money, etc. Using the textual analyses conducted, one can also gain an understanding of the nature of the winning Grants.

References

1. Central Bureau of Statistics. (2021). Gross Domestic Expenditure on R&D. Retrieved from https://www.cbs.gov.il/en/mediarelease/Pages/2021/Gross-Domestic-Expenditure-on-R-D.aspx

2. Israel Strategic Fund. (n.d.). About ISF. Retrieved from https://isf.org.il/about-isf/

3. Israel Innovation Authority. (2017). ISF Grants Program: Performance and Impacts. Retrieved from https://innovationisrael.org.il/en/sites/default/files/ISF%20Grants%20Program%20Performance%20and%20Impacts%20Report%202017%20-%20Final%20Report.pdf

4. Jeroen Baas, Kevin Boyack, John P.A. Ioannidis. (2021). August 2021 data-update for "Updated science-wide author databases of standardized citation indicators". Retrieved from https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/3

5. Grootendorst, Maarten. (2022). "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." arXiv, March 11, 2022. http://arxiv.org/abs/2203.05794

6. Meirmans S, Butlin RK, Charmantier A, Engelstädter J, Groot AT, King KC, Kokko H, Reid JM, Neiman M. (2019). Science policies: How should science funding be allocated? An evolutionary biologists' perspective. *J Evol Biol*, 32(8), 754-768. doi: 10.1111/jeb.13497.

7. Aschhoff, Birgit. (2010). "Who Gets the Money?: The Dynamics of R&D Project Subsidies in Germany." *Jahrbücher für Nationalökonomie und Statistik*, 230(5), 522-546. doi: 10.1515/jbnst-2010-0503

8. Stefano H. Baruffaldi, Marianna Marino, Fabiana Visentin. (2020). "Money to move: The effect on researchers of an international mobility grant." *Research Policy*, 49(8), 104077. doi: 10.1016/j.respol.2020.104077.

9. Jin, Yuping. (2017). "Development of Word Cloud Generator Software Based on Python." In *13th Global Congress on Manufacturing and Management*, 174, 788–792. doi: 10.1016/j.proeng.2017.01.223

10. Parmar, A., Katariya, R., Patel, V. (2019). "A Review on Random Forest: An Ensemble Classifier." In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, Lecture Notes on Data Engineering and Communications Technologies, 26. Springer, Cham. doi: 10.1007/978-3-030-03146-6$_8$6

11. Sheikh, Imran, Irina Illina, Dominique Fohr, Georges Linares. (2016). "Learning Word Importance with the Neural Bag-of-Words Model." In *ACL, Representation Learning for NLP (Repl4NLP) workshop*, Aug 2016, Berlin, Germany. hal-01331720