

# **ACM Kaaryavarta Project**

## **Factors that Affect Entrepreneurship in Indian Colleges**

1st April 2023 | National Institute of Technology Karnataka

Under the guidance of Dr. Rashmi Uchil

Project lead: Swastik Shrey

Project members: Onkar Ray, Nandini Choukimath, Ajinkya Galegave

### **Abstract:**

A detailed analysis of Factors that Affect Entrepreneurship in Indian Colleges. Fostering entrepreneurship among college students depends on the individual's intrinsic personality and external factors. While many studies have focused on the prior, external factors affecting entrepreneurship in India have been left out. This project aims to fill this void by analyzing contextual factors' impacts on college students' entrepreneurial intention.

---

### **Methodology:**

Broadly classified into 3 steps:

- **Literature review** to find a suitable model to determine contextual factors' impact on entrepreneurial intention.
- Conducting a nationwide **survey** for college students with 500+ respondents from 90+ different Indian colleges.
- **Analyzing** the data acquired from the survey to test the model and find insights and limitations of the dataset.

#### **1) Literature Review:**

The process involved defining a research question we'll investigate and then going through the pre-existing literature to choose a relevant model. In our case, we found the **Entrepreneurial Structure Model (ESM)** from the research paper ["Which factors affect the entrepreneurial intention of university students?" by Duygu Turker and Senem Sonmez Selcuk](#) in 2009 to be the most relevant model to answer our research question. We took the ESM model that was tested in Turkey by Dr. Turker and Dr. Selcuk to conduct our research in India.

We did tweak a few questions from the ESM model survey to fit in the Indian equivalent study better.

#### **2) Conducting Survey:**

##### **Questionnaire design**

Our survey was made on a google form with the following questions:

---

**Which college are you from?**

- <Typed Answer>

**Is anyone from your family into Entrepreneurship?**

- Father/Mother/Sibling/Grandparents/No One

**Which year of college are you in currently?**

- First/Second/Third/Fourth/Fifth

*Likert Scale of 1-5 for the other 10 questions.*

*1 -> Not at all*

*5 -> Definitely*

**Educational factors**

**Q1.** *It's easy to find the right mentor in college for my startup.*

**Q2.** *My college and college clubs develop my entrepreneurial skills and abilities.*

**Q3.** *My institution has introduced courses in order to keep us in touch with the growing wave of entrepreneurship in the country.*

**Relational factors**

**Q4.** *My family would support me if I shifted my focus to unconventional or entrepreneurial activities.*

**Q5.** *If I decide to be an entrepreneur, family support will be an important factor for me.*

**Q6.** *I'll readily get mental and executional support from my peers for my venture.*

**Structural factor**

**Q7.** *It is easy for me to get funding for my startup from alumni or outside.*

**Q8.** *My country has created an ecosystem that enables young entrepreneurs to pursue their passion.*

**Q9.** *My college is located in a city which is active in entrepreneurship.*

**How likely are you to take up entrepreneurship as a career?**

---

All these questions have been taken directly from or modified slightly from the ESM model. The 3 factors that are being tested in the ESM model are **Perceived Educational Support (PES)**, **Perceived Structural Support (PSS)**, and **Perceived Relational Support (PRS)**, which are being tested against the Perceived Entrepreneurial intention of College students, which is measured by the response to the final question.

- Questions 1, 2, and 3 to test the Educational Factor
- Questions 4, 5, and 6 to test the Relational Factor
- Questions 7, 8, and 9 to test the Structural Factor

Similar to the ESM model in the Toker 2009 research paper.

Apart from these questions, we have also taken basic information about the respondents (anonymous), such as **college name**, **year of study**, and **family's entrepreneurial history**, which help us draw insights about our sample and find our limitations.

Our survey recorded **501 responses** at the time when the data was compiled and used for analysis.

### 3) Analysis:

The software and resources we used for our analysis included:

- **Microsoft Excel**
- **Python (libraries: NumPy, Pandas, Matplotlib, Seaborn, Statistics)**
- **ChatGPT (3.0/4.0)**
- **Plotly Chart studio**

We obtained the data in .xlsx file format from the google form response survey. We then **cleaned the data** and worked on solving the problem of people entering different abbreviations, names, and upper/lower case mismatch for the same college.

- We **assigned a key to each college** and created a new column in excel that identified colleges through their key irrespective of the naming and casing anomalies. We did this with the help of GPT 3.0, which was instructed to read through the list of colleges, identify which college it was and give an output of a table with a common college name and key.
- **New columns containing the difference** between answers to each of the 9 questions, along with the mean response for each major factor, Educational, Structural and Relational, and Perceived Entrepreneurial Intention, were created to be used for analysis.
- Finally, we added two new columns, namely, the **"City"** in which the college was located and the type of college, **"Public or Private"**, to get more insights about the sample data.

With the data about the city of respondents, a **Density Tile Map of India** was created using Plotly Chart Studio to visualize the density and spread of respondents from different parts of India.

Further analysis was done using Python and its libraries. We uploaded the CSV files from excel to GitHub and imported them to our ipynb (python notebook) along with the necessary libraries. Here are the further steps we proceeded with:

- Creating a new DataFrame with required columns
- **Exploratory Data Analysis**
- Establishing method of correlation: Using Distribution curve and Standard deviation (devised a new way of correlating data of this type. Methodology explained after this)
- **Evaluated correlation** for all 9 Questions along with the 3 major factors of correlation
- Finding perceived entrepreneurial interest, funding, relational support, etc, for different cities, colleges, years of education, etc, and **ranking** them using bar plots to visualize.
- Finding a few more **insights and limitations** using pie charts and distribution curves.
- **Beautifying visualizations** in seaborn for creating the report.

---

## Method of correlation:

Pandas' inbuilt correlation functions `.corr()` do not work well for discrete and low-range datasets like ours. Hence we are using a unique way of establishing a correlation between the factors and entrepreneurship

***Note: Every factor and entrepreneurship in this report is as "perceived" by the respondents.***

For example: To find the correlation between perceived entrepreneurship(Ent) and perceived family support(Q4)

- We find the difference between the response for Q4 and Ent for each response
- In an idealistic direct correlation, the difference should be 0 for all respondents, but we see divergence for some responses.
- In simple terms, if the difference is high, then the correlation is less; hence we plot a distribution graph of this difference.
- Difference  $\leq |1|$  implies some correlation whereas difference  $\geq |2|$  points towards a weak or no correlation
- Hence, the distribution curve's spread helps us understand the overall correlation of the factor. The less distributed the curve, the higher the correlation.
- We can also mathematically describe less distribution as the standard deviation of the distribution curve being low. Hence, the lower the standard deviation, the higher the correlation.

**A standard deviation of 1.33 would occur for a normal distribution curve in our case. Hence a factor with a standard deviation of more than 1.33 shows a weaker correlation with perceived entrepreneurship.**

---

## Findings:

The result of the survey analysis showed that in India, the **Perceived Relational Factor (PRS)** had the highest correlation with **Perceived Entrepreneurial Interest in Indian College Students**. Followed by **Perceived Structural Support (PSS)** and then **Perceived Educational Support (PES)**. This also demonstrated the difference in culture between Turkish and Indian societies as the Turkish ESM research concluded that Perceived Educational Support (PES) had the highest correlation, followed by Perceived Structural Support (PSS) and then Perceived Relational Factor (PRS).

---

## Practical Implications:

The study might be of help to the **policymakers in the Indian Education system in order to create a curriculum and environment in colleges that aid entrepreneurial interest.**

---

## Limitations:

The study has a few limitations that can be overcome with a more comprehensive study with a better sample set. A few of the limitations that can be overcome are:

- **Testing only a few factors per the ESM model** ignores some important factors from other studies, such as Reduced Risk Factors - Monetary or due to government aid playing a huge role in Entrepreneurship, as found in studies done in the US and Canada.
- **Self-selecting sample bias:** More entrepreneurially inclined students responding to the data as found out during analysis.
- **Convergence bias:** Getting more responses from Engineering students, 2nd-year students, and some colleges being represented by just one respondent.
- 

Limitations that cannot be overcome include all the answers being the perception of the respondent about their college, city, etc. Even their interest in entrepreneurship is perceived.

Testing the analysis results with some data helps establish that the outcome is reliable to a certain level. Comparing the analysis of the colleges with the highest number of startups, which receive the highest funding, and the cities which are the most active in entrepreneurship are some metrics that were tested against the actual available data.

---