

# University Major Recommendation System

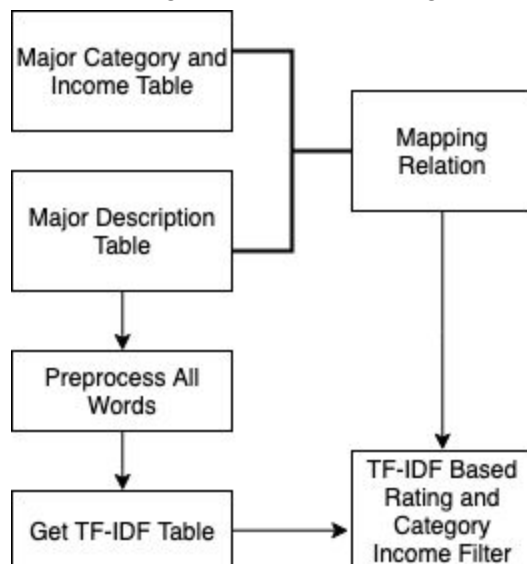
## Motivation and Problem Understanding

Choosing a major is one of the most important decisions people make in their lives. The right major affect people's future in a way that you can never imagine. According to the U.S. Department of Labor, the average twenty-something switches jobs once every three years and the average person changes career fields two or three times in their lifetime.

A major recommendation system will be a good tool to search the database in giving recommendations on majors that fit the need of the students based on their requirements and also personalized to fit their interests and profile. As a result, this system's goal is to reduce the time spent in searching for majors for students who need to declare a major or simply to find interest fields to dive into. We also aim to recommend the right major to the user which suits their capability and interest and potentially saving the cost of reinvest in study after college education.

## Original Data Processing

The following chart is the working flow for the original data processing:



## Major Description Data Processing

Used a web crawler to download the original text data contains all the major's names and their descriptions. Stored the original data to a file with special characters to separate each major, in order to conveniently read the data to memory. Read the major description from the file, use a map to store all the majors, the key is the major name and the value is the description. Change all the words in the map, from the last step, to lowercase, delete special characters (like ,.\*!), and delete all the non-important words (like a, an, the, in). Created the final major description map whose key is the major name and value is the list of words from the original description.

For each word in the map from the last step, if it is a plural word (end with 's' or 'es', 'ies'), add the original one to the description word list, e.g., for "sciences", add "science" to the list. This step is necessary as the original version should have the same or higher weight than the plural version, and user's input could contain non-plural words. Create a set which contains all the unique words, based on which can get the TF-IDF value vector for each major.

Get the TF-IDF values. For each major, calculate the TF-IDF value of each word in the description word list. Allocate weights to words in the major name, the weight is the name for each word in the major name and the value equals to two times the maximum TF-IDF value among the description words, since major name should have higher weight for a match. Finally get a map in which the key is the major name, and value is the TF-IDF value vector.

## Other Major Info Processing

Downloaded the data contains other info about all the majors like, its category, graduation rate, and medium income after graduation. Since this data is from another source, the major name is different from the one in the major description info. Therefore, I manually compare those two versions of major names and create a mapping between two major names in different version and store it in a text file, in which the key is the major name in this table and the value is the major name in the major description map. Create a map in which the key is the major name, the value contains its category, graduation rate and medium income. Change all the words in the map to lower case, delete non-important characters.

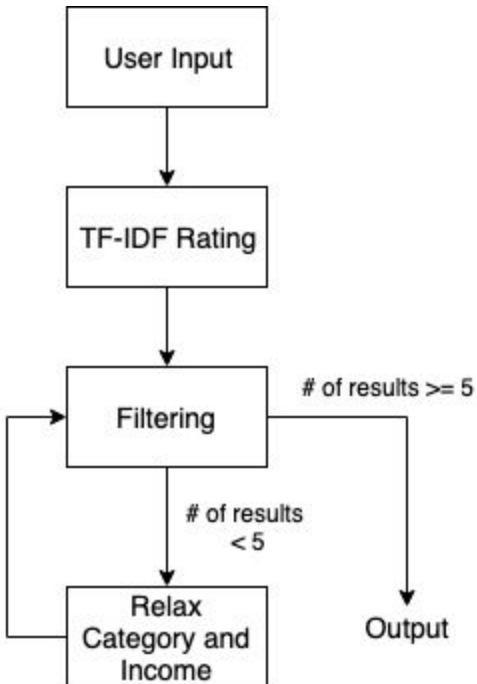
## User Input Processing

User's input is a string contains the information about the user's preference towards a major. First split the string to a word list, delete useless characters in each word, and exclude non-important words in the list (like a, an, the).

Assign weights to the input words. Assign weights to words based on the order of each word, i.e., weight decreases as order increases, since a user prefer to input more preferred words in the description first. Use `numpy.linspace(start=10, end=5, count)` to allocate weights. The end is 5 by default not 0 is because words appear later in the input could also be important, the weights for them should be conservative.

## Recommendation Algorithm

The user will input three inputs: user's requirement or description of the preferred major, the minimum income (optional), select a major category from UI (optional). The following is the flow chart for the summary of the algorithm:



Based on the already addressed data processing, the input major description will be transferred to a list of words. Next is to calculate the initial rating for each major. The calculation is very simple, just sum the tf-idf value of the description word in the major's tf-idf value map. After getting the rating for each major, sort all the majors based on the ratings to get the initial major ranking. In this application, we statically set the number of recommended majors to the user is 5, which is reasonable and practical that too many recommendations will confuse the user.

The next step is to incorporate the major category and minimum income constraints to get more precise recommendations based on the user's preference. After adding constraints, a practical problem is that there will be no equal or more than 5 results, under which scenario we need to relax the constraints to make the recommended results be at least 5. Since there two constraints, the major category and minimum income, a relaxing order or a relaxing priority should be assigned to the two ones. In this application, we will first relax the major category to none, since major category is a single-value constraint. After delete the major category if there are still less than 5 results, then each round the algorithm will decrease the minimum income constraint by \$5000, until we finally find 5 satisfied

recommendations. During the relaxing process, we still use the initial major ranking based on the tf-idf value. Therefore, the final results will still be sorted by ratings.

## User Facing Application

The user will input three inputs: user's requirement or description of the preferred major, the minimum income (optional), select a major category from UI (optional).

### College Major Recommendation System

Describe your advantages and expectations for the major (Required\*)

Minimum Expected Annual Salary After College Graduation(Optional)

eg, 50000

Please select at most one preferred category(Optional)

Choose...



GET RECOMMENDED MAJORS

### Case 1: With description and minimum income.

Describe your advantages and expectations for the major (Required\*)

I like mathematics and physics a lot, want to do engineering works

Minimum Expected Annual Salary After College Graduation(Optional)

50000

Please select at most one preferred category(Optional)

Choose...



GET RECOMMENDED MAJORS

College Major Recommendation System				
name	rating	category	income	Info
Nuclear Engineering	755.1674161485926	Engineering	95000.0	Nuclear engineering, the practical application of the principles of nuclear science for the benefit of humankind, provides the engineer or scientist with an opportunity to work on challenging problems that are vitally important to the modern world. The goal of many nuclear engineering programs is to provide students with a strong academic background that enables them to pursue professional careers in nuclear and radiation-based industries, or to pursue graduate study in nuclear engineering or related fields such as medical physics, health physics, or another field of engineering. Graduates may be hired by electric power companies that use nuclear power plants, or by companies that help service and maintain those plants. They use their knowledge of engineering principles, radioactive decay, interactions of radiation with matter, and nuclear reactor behavior to help assure that the power plants meet the demand for reliable, economic electricity while ensuring a safe environment. Other graduates work in industries that use radioactivity or radiation to detect problems or monitor processes. Jobs are also found in branches of the government as designers of the next generation of reactors for submarines, aircraft carriers, or space probes, or to manage and clean up contaminated wastes. They could also be involved with regulation of nuclear power or radiation uses, or in research to develop advanced technologies that will be used in next-generation power plants. Graduates who want to further their education in the fields of health physics, radiation biology, or nuclear medical applications find this degree to be a useful preparation. [ENGINEERING]
Mathematics	438.559095577133	Computers & Mathematics	66000.0	Mathematics today is a vibrant and dynamic field, manifesting itself in such familiar things as CAT scans, compact discs, and satellite communication. Accordingly, the mathematics program is broad and varied, with studies in calculus, finite, actuarial sciences and algebra. After graduating, majors successfully pursue a variety of careers, working in industry, insurance (as actuaries), government, finance, and other fields. Many graduates may attend professional schools in law, medicine, or business. Others may go directly into teaching; and still others may go on to graduate programs in mathematics, applied mathematics, engineering, economics, and computer science. [MATHEMATICAL SCIENCES]
Engineering Science	401.3129950954393	Engineering	63000.0	In Engineering Science understanding of the basic engineering practices are combined with an in-depth knowledge of the fundamentals of the engineering sciences: electricity and magnetism, fluid and solid mechanics, mathematics, computer applications, materials, and thermodynamics and heat transfer. Graduates may choose to pursue graduate study either in biomedical engineering or medicine, or a career in one of the traditional engineering disciplines. [MATHEMATICAL SCIENCES/ENGINEERING/ SCIENCE]
Environmental Systems Engineering	401.3129950954393	Engineering	70000.0	Environmental Systems Engineering is an interdisciplinary program concerned with the sources and causes of industrial impact on the environment and the choice of appropriate remediation strategies. Students enrolled in the program acquire: a) a strong foundation in mathematics and the physical and engineering sciences, b) a fundamental background in the application of the earth and atmospheric sciences and the principles of process engineering to the environmental problems of the basic industries, especially those involved in the extraction, conversion and utilization of minerals and fuels, c) the ability to carry out experiments, interpret data and apply the results to the design and operation of engineering systems, and d) an understanding of the human, societal and economic issues involved in industrial activity and environmental protection. [ENGINEERING]
Physics	400.49355947431314	Physical Sciences	70000.0	Physics studies the fundamental properties and interactions of all forms of matter. Experimental and theoretical investigations are combined to formulate mathematical relationships that describe and predict the behavior of nature. The physics undergraduate program can prepare students for employment in an industrial or governmental laboratory. The program can also prepare students for further study at graduate or professional schools in physics, engineering, biophysics, medicine, education, law, or business. [MATHEMATICAL SCIENCES/ SCIENCE]

The result is reasonable since nuclear engineering which combines engineering, physics, and mathematics. We can also notice that all the other four ratings are very close since they only covers part of the user's major description.

## Case 2: With description, expected salary = 50, 000 USD, no category selected.

Describe your advantages and expectations for the major

I like reading naval and like writing stories and articles.

Minimum Expected Annual Salary After College Graduation(\$)

50000

Please select at most one preferred category(Optional)

Choose...

GET RECOMMENDED MAJORS

Result:

College Major Recommendation System				
name	rating	category	income	Info
Geography	122.8771237954945	Social Science	54000.0	Geographers describe, analyze, and explain patterns of physical and human phenomena on the Earth's surface. Geography is simultaneously an environmental science and a social-behavioral science. This major helps provide grounding in analytical techniques such as map reading, cartography, and statistics. Substantive course work investigates the ways people use environmental resources and how they arrange themselves and their economic, social, and political activities on the earth's surface. The Geography major can provide preparation for a career in business, industry, or government. Geographers with bachelor's degrees may find placement in federal, state, and local administrative and planning agencies or in private firms that specialize in planning and development or in environmental and socioeconomic analysis. [SOCIAL SCIENCE]
English	101.17787378107137	Humanities & Liberal Arts	50000.0	This major provides an opportunity to study human communication and artistic expression through literature language, writing, and theory. English majors learn to write clearly and effectively, think critically and analytically, and question the works and the world around them. While these skills are applicable to virtually any future career choice, most English majors apply to graduate school, law school and even medical school. Those who do not pursue graduate school find the study of English an excellent preparation for government service, business careers, international agencies, publishing, journalism, and secondary school teaching. [COMMUNICATION ARTS]
Journalism	101.17787378107137	Communications & Journalism	50000.0	Careers include newspaper reporting and editing, magazine writing, broadcast news, corporate public relations, and advertising account services. This major is based on knowledge of the social and professional responsibilities of communicators, and basic competence in journalistic skills. [COMMUNICATION ARTS]
Accounting	0.0	Business	65000.0	Accounting supplies quantitative information essential to management decision-making and control, as well as a wide variety of tax and consulting services and information on management's effective use of an organization's resources. This major helps prepare students for careers in public, industrial, or governmental accounting and also provides an appropriate background for those planning to enter law school or graduate school. Public accounting is carried on by independent practitioners, most of whom are certified public accountants. [BUSINESS]
Actuarial Science	0.0	Business	72000.0	An actuary is a business professional who uses mathematical skills to define, analyze, and solve business and social problems. Actuaries generally are employed in life, health, and casualty insurance companies, consulting firms, and government. [MATHEMATICAL SCIENCES]

The top three results are pretty close as they all match the description. All the other ratings are zero so two random majors are returned.



## Case 3: Same description and salary, category choose “Communication & Journalism”

Describe your advantages and expectations for the major (Required\*)

I like reading naval and like writing stories and articles.

Minimum Expected Annual Salary After College Graduation(Optional)

50000

Please select at most one preferred category(Optional)

Communications & Journalism

GET RECOMMENDED MAJORS

Output:

College Major Recommendation System				
name	rating	category	income	Info
Geography	122.8771237954945	Social Science	54000.0	Geographers describe, analyze, and explain patterns of physical and human phenomena on the Earth's surface. Geography is simultaneously an environmental science and a social-behavioral science. This major helps provide grounding in analytical techniques such as map reading, cartography, and statistics. Substantive course work investigates the ways people use environmental resources and how they arrange themselves and their economic, social, and political activities on the earth's surface. The Geography major can provide preparation for a career in business, industry, or government. Geographers with bachelor's degrees may find placement in federal, state, and local administrative and planning agencies or in private firms that specialize in planning and development or in environmental and socioeconomic analysis. [SOCIAL SCIENCE]
English	101.17787378107137	Humanities & Liberal Arts	50000.0	This major provides an opportunity to study human communication and artistic expression through literature language, writing, and theory. English majors learn to write clearly and effectively, think critically and analytically, and question the works and the world around them. While these skills are applicable to virtually any future career choice, most English majors apply to graduate school, law school and even medical school. Those who do not pursue graduate school find the study of English an excellent preparation for government service, business careers, international agencies, publishing, journalism, and secondary school teaching. [COMMUNICATION ARTS]
Journalism	101.17787378107137	Communications & Journalism	50000.0	Careers include newspaper reporting and editing, magazine writing, broadcast news, corporate public relations, and advertising account services. This major is based on knowledge of the social and professional responsibilities of communicators, and basic competence in journalistic skills. [COMMUNICATION ARTS]
Accounting	0.0	Business	65000.0	Accounting supplies quantitative information essential to management decision-making and control, as well as a wide variety of tax and consulting services and information on management's effective use of an organization's resources. This major helps prepare students for careers in public, industrial, or governmental accounting and also provides an appropriate background for those planning to enter law school or graduate school. Public accounting is carried on by independent practitioners, most of whom are certified public accountants. [BUSINESS]
Actuarial Science	0.0	Business	72000.0	An actuary is a business professional who uses mathematical skills to define, analyze, and solve business and social problems. Actuaries generally are employed in life, health, and casualty insurance companies, consulting firms, and government. [MATHEMATICAL SCIENCES]

Choosing category did not change the result since there are still three majors has non-zero ratings and the category constraint will be relaxed during the first round of relaxing process, since in the algorithm we will first relax the category constrain.