**Research Work on Venture Capital**

**Private Equity-**

Private equity typically refers to investment funds, generally organized as limited partnerships, that buy and restructure companies that are not publicly traded.

Private equity is composed of funds and investors that directly invest in private companies, or that engage in buyouts of public companies, resulting in the delisting of public equity. Institutional and retail investors provide the capital for private equity, and the capital can be utilized to fund new technology, make acquisitions, expand working capital, and to bolster and solidify a balance sheet.

A **private equity fund has Limited Partners (LP),** who typically own 99 percent of shares in a fund and have limited liability, and **General Partners (GP)**, who own **1 percent of shares** and **have full liability**. The latter are also responsible for executing and operating the investment.

**General Partnerships**

A general partnership is the most common type of partnership. It refers to a relationship in which all partners contribute to the day-to-day management of the business. Each partner will have the authority to make business decisions and even legally bind the company in contracts.

The liabilities, contributions, and responsibilities of the partners are often equal unless stated otherwise. Typically, a partnership agreement will describe which partners have certain authorities and responsibilities.

**Limited Partnerships**

A limited partnership is a relationship where one or more partners are not involved in the day-to-day management of the business. Often, a limited partner, sometimes known as a “silent partner,” will serve solely as an investor in the business, with the funds that they contribute being the extent of their liability. However, since the limited partner does not have decision-making power in the company, withdrawing funds – even just the amount they’ve already contributed – cannot be done without the approval of a general partner. Limited partnerships will still have at least one general partner to man the day-to-day operations of the business. A general partner may invest money into the company. However, a general partner may also be personally liable for the debts of the company, while the limited partner is not. Only a general partner’s personal assets (in addition to the business assets) can come into play when it comes to paying off the company’s debts.

**NLP Research**

Tokenization – it is just grouping or words or sentence tokenizer separates by sentences

**Lexicon & Corpora’s**

**Corpra –** Body of text . ex: medical journal , presidential speeches

Lexicon – words and meaning . Or difference investor speak vs regular speak

For ex. If an investor uses “bull ” it might mean market is good , but regularly it might mean a scary animal.

Text Preprocessing –

1. ***Stop Words*** - Sentiment of words which you don’t know – ‘stop words or words, Words which you don’t need them – eg , a , the etc … Basically called as filler words
2. ***Stemming*** – Ride – riding, ridden , rides . We do it because , meaning of the words would be unchanged .[which basically takes up spaces ]. We use steming to deal with this. Prter stemmer[1995]
3. ***Part of speech tagging*** – Assiging part of speech to a single word
4. ***Chunking*** - Next step fig. out sentence what it is talking about. We we many noun in a sentence. It contains **modifiers** which is basically starting of a sentence ($) replaced by symbols. (Explanation- If you want to find out how US president says his speech- Noun + “some proper noun” + verb …..etc )
5. ***Chinking***- Removal of something. Maybe if you want to remove what has been chunked.
6. ***NER*** – **Named Entity Recognition**

**What does tf-idf mean? (which will be used for finding important keywords)**

Tf-idf stands for term frequency-inverse document frequency, and the tf-idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. Variations of the tf-idf weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query.

One of the simplest ranking functions is computed by summing the tf-idf for each query term; many more sophisticated ranking functions are variants of this simple model.

Tf-idf can be successfully used for stop-words filtering in various subject fields including text summarization and classification.

**How to Compute:**

Typically, the tf-idf weight is composed by two terms: the first computes the normalized Term Frequency (TF), aka. the number of times a word appears in a document, divided by the total number of words in that document; the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.

**TF**: Term Frequency, which measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (aka. the total number of terms in the document) as a way of normalization:

TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).

**IDF**: Inverse Document Frequency, which measures how important a term is. While computing TF, all terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

IDF(t) = log\_e(Total number of documents / Number of documents with term t in it).

See below for a simple example.

Example:

Consider a document containing 100 words wherein the word cat appears 3 times. The term frequency (i.e., tf) for cat is then (3 / 100) = 0.03. Now, assume we have 10 million documents and the word cat appears in one thousands of these. Then, the inverse document frequency (i.e., idf) is calculated as log(10,000,000 / 1,000) = 4. Thus, the Tf-idf weight is the product of these quantities: 0.03 \* 4 = 0.12.