A Comprehensive and Subjective Job Search Recommendation System

Presenter: Di Liu
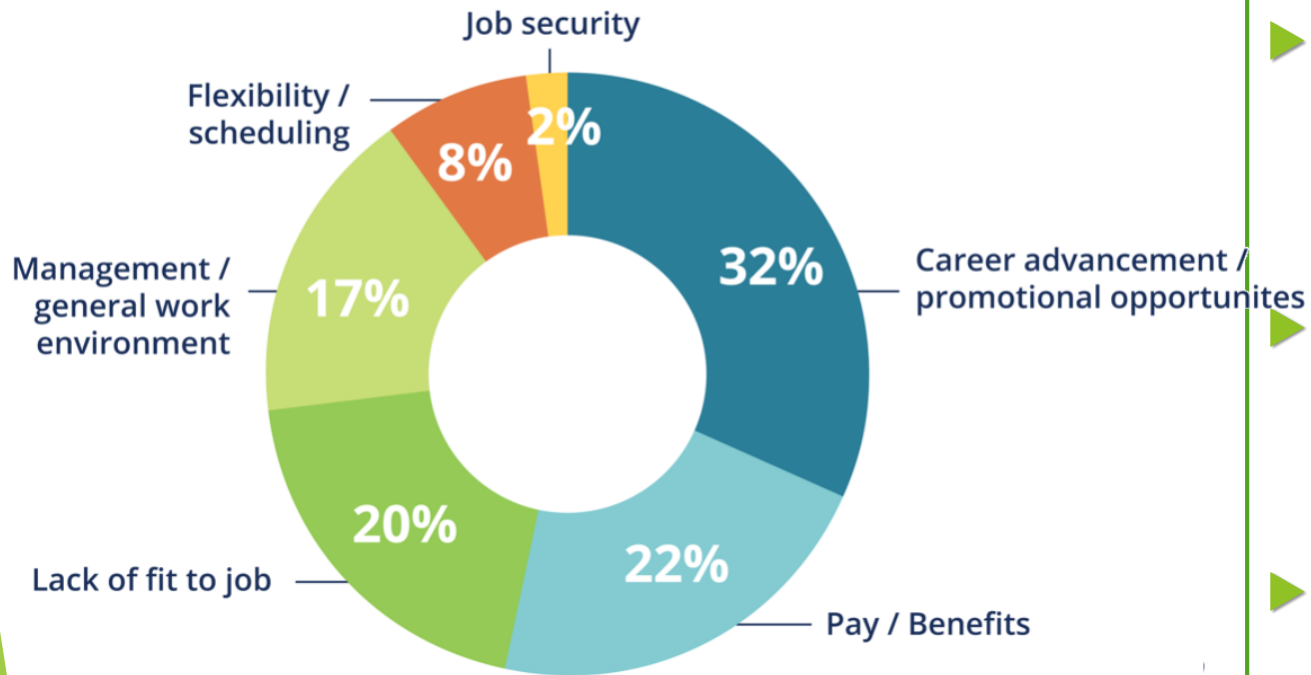
Ph.D. in Mechanical Engineering

Virginia Commonwealth University, Richmond, Virginia

2019/2/14

Motivation

TOP VOLUNTARY JOB-RELATED REASONS FOR CHANGING JOBS

Job security
Flexibility / scheduling
Management / general work environment
Lack of fit to job
Career advancement / promotional opportunites
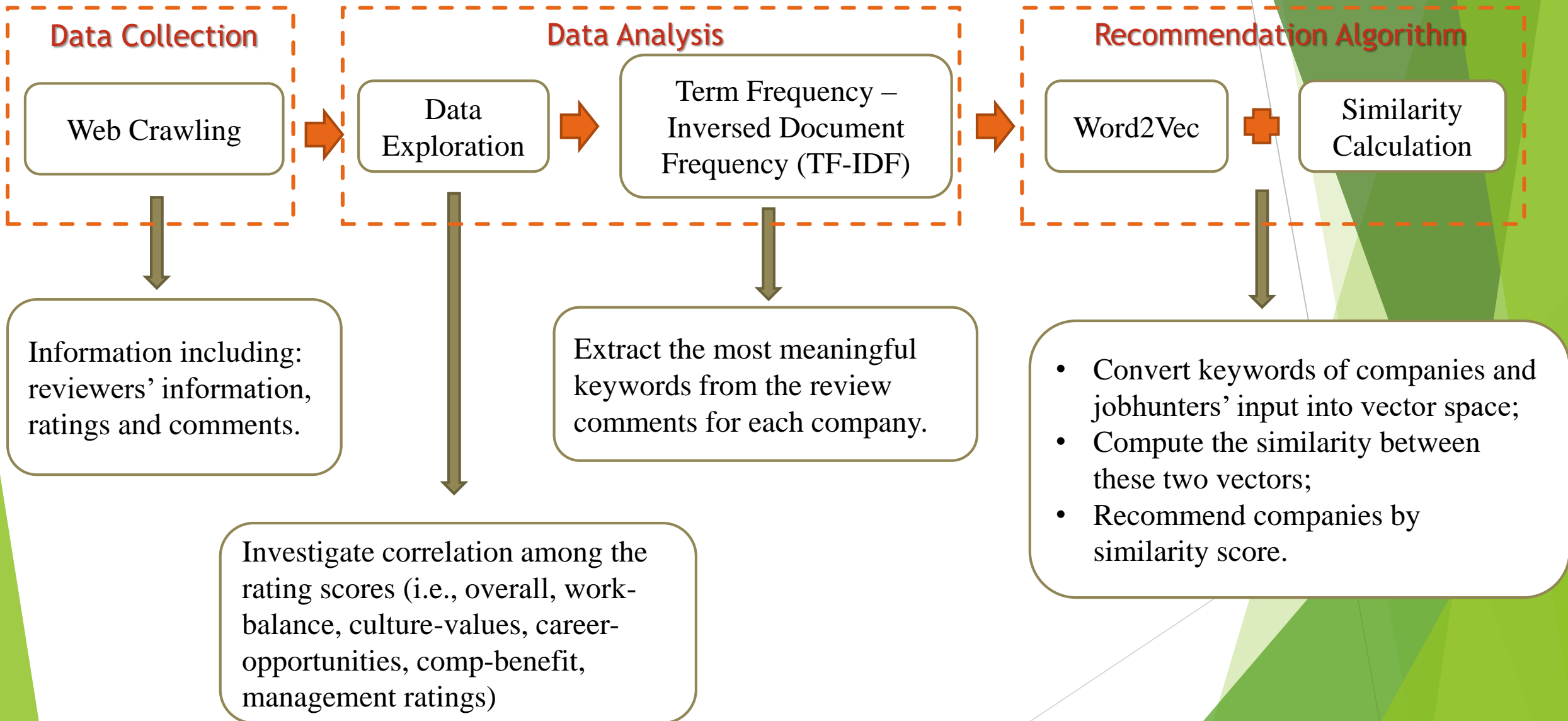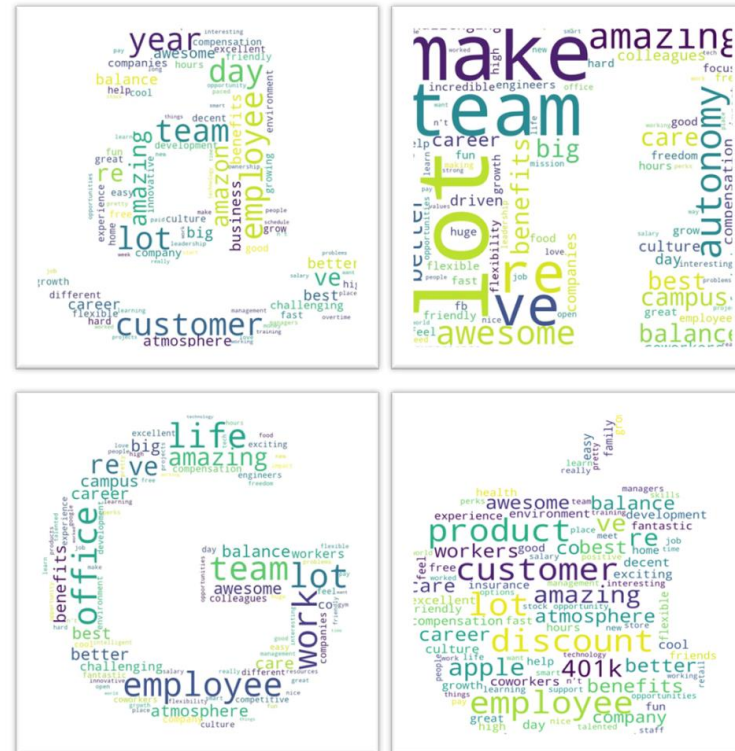Pay / Benefits

2%
8%
17%
20%
32%
22%

Source: Gallup
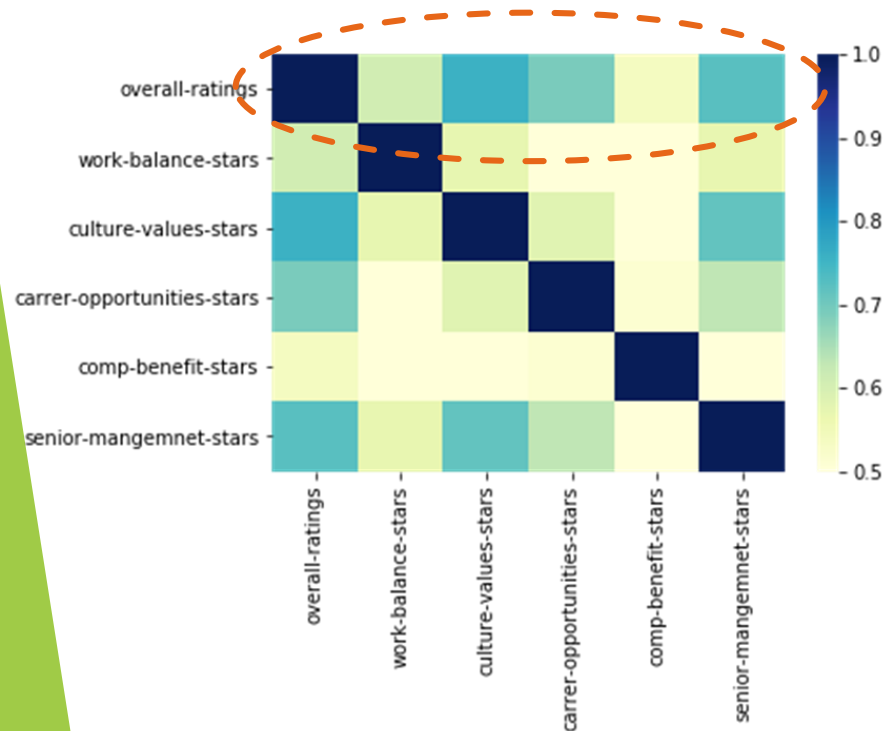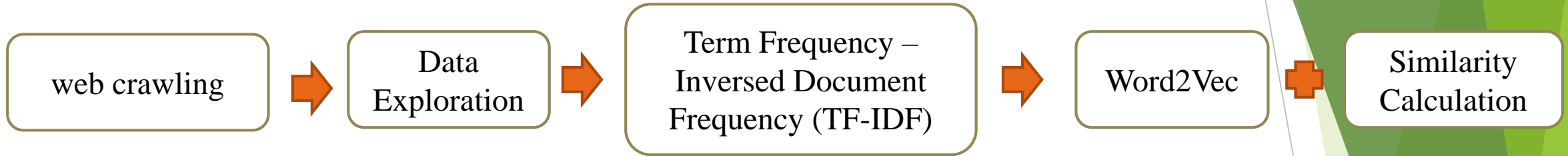(http://myaliat.com/news/employees-mentally-checked-out/top-voluntary-job-related-reasons-changing-jobs-gallup-aliat/)

▶ **Frequently changing jobs has adverse effects on both employers and employees themselves.**

▶ **Unlike 'Pay/benefit', about 80% reason for people changing jobs are not quantifiable.**

▶ **Comprehensive and subjective understanding of the companies during the job searching process is important.**

System Design

Data Collection

Web Crawling

Information including: reviewers' information, ratings and comments.

Data Analysis

Data Exploration

Term Frequency – Inversed Document Frequency (TF-IDF)

Investigate correlation among the rating scores (i.e., overall, work-balance, culture-values, career-opportunities, comp-benefit, management ratings)

Extract the most meaningful keywords from the review comments for each company.

Recommendation Algorithm

Word2Vec + Similarity Calculation

- Convert keywords of companies and jobhunters' input into vector space;
- Compute the similarity between these two vectors;
- Recommend companies by similarity score.

Takeaway so far…

web crawling → Data Exploration → Term Frequency – Inversed Document Frequency (TF-IDF) → Word2Vec ✚ Similarity Calculation

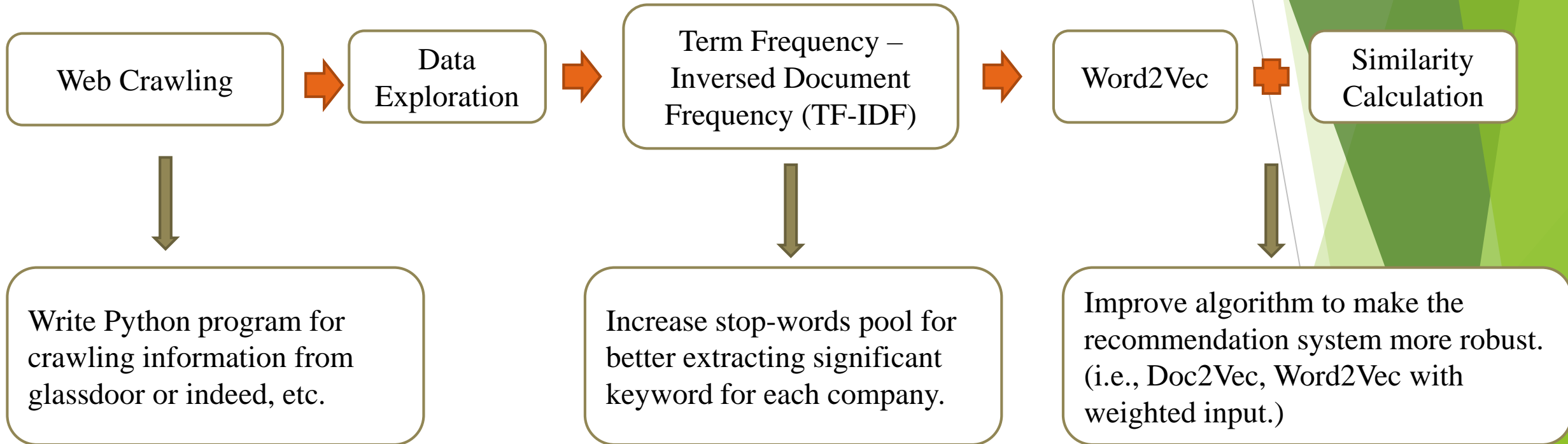

Darker color represents that the two factors are more co-related.

➢ **The overall ratings are more related to the company culture and management rather than 'money'.**

➢ **Wordclouds including the keywords have been plotted for companies.**

To Do List

| Web Crawling | → | Data Exploration | → | Term Frequency – Inversed Document Frequency (TF-IDF) | → | Word2Vec | ➕ | Similarity Calculation |

Write Python program for crawling information from glassdoor or indeed, etc.

Increase stop-words pool for better extracting significant keyword for each company.

Improve algorithm to make the recommendation system more robust. (i.e., Doc2Vec, Word2Vec with weighted input.)

*There is no **Mr.Wrong**,*

*but **Mr.Right** in the **WRONG** place!*

Thank you!
& Question?

* For more information, please refer to my github link: https://github.com/liudi1025/TDI