Project 4:
A job-role recommender system based on (introverted) MBTI personality traits

Brought to you by **THE BURNT-OUT BUNNIES**

#happyemployee
#happyboss

#SG-DSI-41

Constance, Germaine & Matthew

Unemployment in Singapore

1. In targeting to reduce unemployment in Singapore, mid-career job-seekers form a main proportion of unemployment numbers
   a. 63% of job seekers over the age of 45 are unemployed for over a year, compared to only 36% of job seekers aged 18 to 24.
2. Even after job-seekers are employed, ensuring employee retention is increasingly challenging
   a. Inefficient resource use across all stakeholders (job seeker, employers, Government) with lower retention rates
   b. Job satisfaction plays a major role in ensuring that employees stay longer

Evidences

Burnout among top business risks predicted for 2024

https://www.hcamag.com/asia/specialisation/corporate-wellness/burnout-among-top-business-risks-predicted-for-2024/469748

Less Freedom, More Burnout: Work Life Needs Help In 2024

Can I embrace my work quirks without feeling like a failure?

Adrianna Lakatos · Follow
11 min read · Jan 6

We aren't exactly thriving — and I know I'm not alone in feeling this way.

Is the answer sending us all back to the office? Or do we need to take time to reconsider our relationship with work?

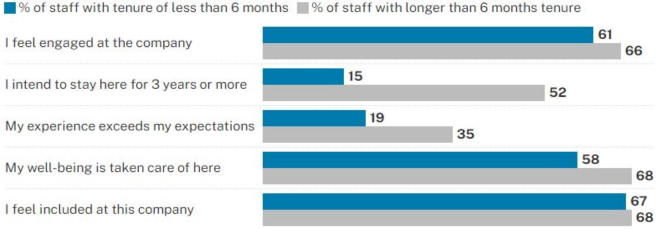
I seem to always be working "wrong".
I can't seem to force myself neatly into any one box.
I don't think my brain works how society expects it to.
I feel weird calling myself an entrepreneur, but I'm not cut out for a 9–5.
I work in strange ways, but if not for comparison with others, I wouldn't see anything wrong with my quirks.

https://medium.com/@adriannalakatos/less-freedom-more-burnout-work-life-needs-help-in-2024-62e2f9cab23a

The 'new job honeymoon' phase is over for workers in Singapore

BT
Cecelia Herbert
Published Mon, Nov 20, 2023 · 5:00 am

Most new joiners don't plan to stay more than 3 years



SOURCE: QUALTRICS
GRAPHIC: BTVISUAL

https://www.businesstimes.com.sg/working-life/new-job-honeymoon-phase-over-workers-singapore

Unemployment in Singapore

1. In targeting to reduce unemployment in Singapore, mid-career job-seekers form a main proportion of unemployment numbers
    a. 63% of job seekers over the age of 45 are unemployed for over a year, compared to only 36% of job seekers aged 18 to 24.
2. Even after job-seekers are employed, ensuring employee retention is increasingly challenging
    a. Inefficient resource use across all stakeholders (job seeker, employers, Government) with lower retention rates
    b. Job satisfaction with successful jon-fits plays a major role in ensuring that employees stay longer

Reduce unemployment by ensuring mid-career job seekers are quickly employed/ upskilled, and continue to stay in their new role with increased job satisfaction

Personality traits in successful job-fits

Utilising personality traits indicators to identify suitable job roles for improved job satisfaction in job seekers

JOB CHARACTERISTICS AND PERSONALITY AS PREDICTORS OF JOB SATISFACTION

Adrian Thomas, Walter C. Buboltz, Christopher S. Winkelspecht

Organizational Analysis
ISSN: 1551-7470
Article publication date: 1 February 2004 Permissions

DOWNLOADS
4057

https://www.emerald.com/insight/content/doi/10.1108/eb028993/full/html

THESIS | OPEN ACCESS

The effect of personality on team boosting behaviours

Jennifer Mmatlou Rammutla

MCom, University of Johannesburg
2023
Handle: https://hdl.handle.net/10210/505361

View Share Export

https://ujcontent.uj.ac.za/esploro/outputs/graduate/The-effect-of-personality-on-team/9934309207691

Looking Ahead, Looking Around, and Looking to Others: Identifying Core Proactive Behaviors in the Quest for Career Sustainability

Robert W. Lent, Steven D. Brown, [...], and Bhanu Priya Moturu +2 View all authors and affiliations

OnlineFirst | https://doi.org/10.1177/10690727231209777

https://journals.sagepub.com/doi/abs/10.1177/10690727231209777

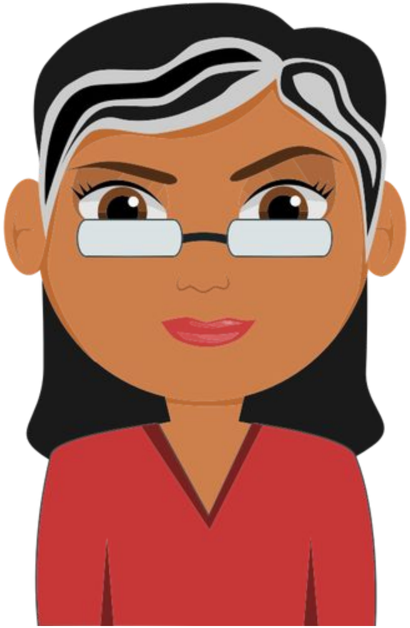
Background of MBTI

The **Myers-Briggs Type Indicator (MBTI)** is a personality type measure that focuses on 4 different categories:

1. Energy (**E**xtrovert/ **I**ntrovert)
    - Energy is the scale of extraversion to introversion and how they direct their attention and how they derive energy i.e. from their surroundings or from solitude.
2. Perceiving (**S**ensing/ I**n**tuitive)
    - Perceiving is the preference of intaking information such as using all 5 senses or based on intuition.
3. Judging (**T**hinking/ **F**eeling)
    - Judging is to categorize how one makes decision, either basing it on logic and facts vs the method of solving an issue such as through harmonious team dynamics.
4. Orientation (**J**udgement/ **P**erceiving).
    - Orientation is about a preference of orderly and decisive lifestyle or a more flexible type of lifestyle.

Source: Myers Brigg

Sharon Tan



Background:

- 42 year old, Singaporean, **family of 3 kids**
- Recently resigned from a healthcare professional role that directly interfaces with patients
- Feeling completely **dissatisfied with previous role** - does not enjoy talking to others, role is routine and lacks challenge
- **Unfamiliar with concept of personality trait indicators e.g. MBTI**

Motivations/ Goals

- Growth mindset, desires to upskill for her next role
- Reserved and do not enjoy small talk
- Obsessed with structure and numbers, thrives on logic and reasoning

Frustrations/ Pain points:

- Desires to enter into tech industry, but **uncertain of the exact roles suited for her**
- **Unsure and very selective on courses** to take for upskill; due to tight budget

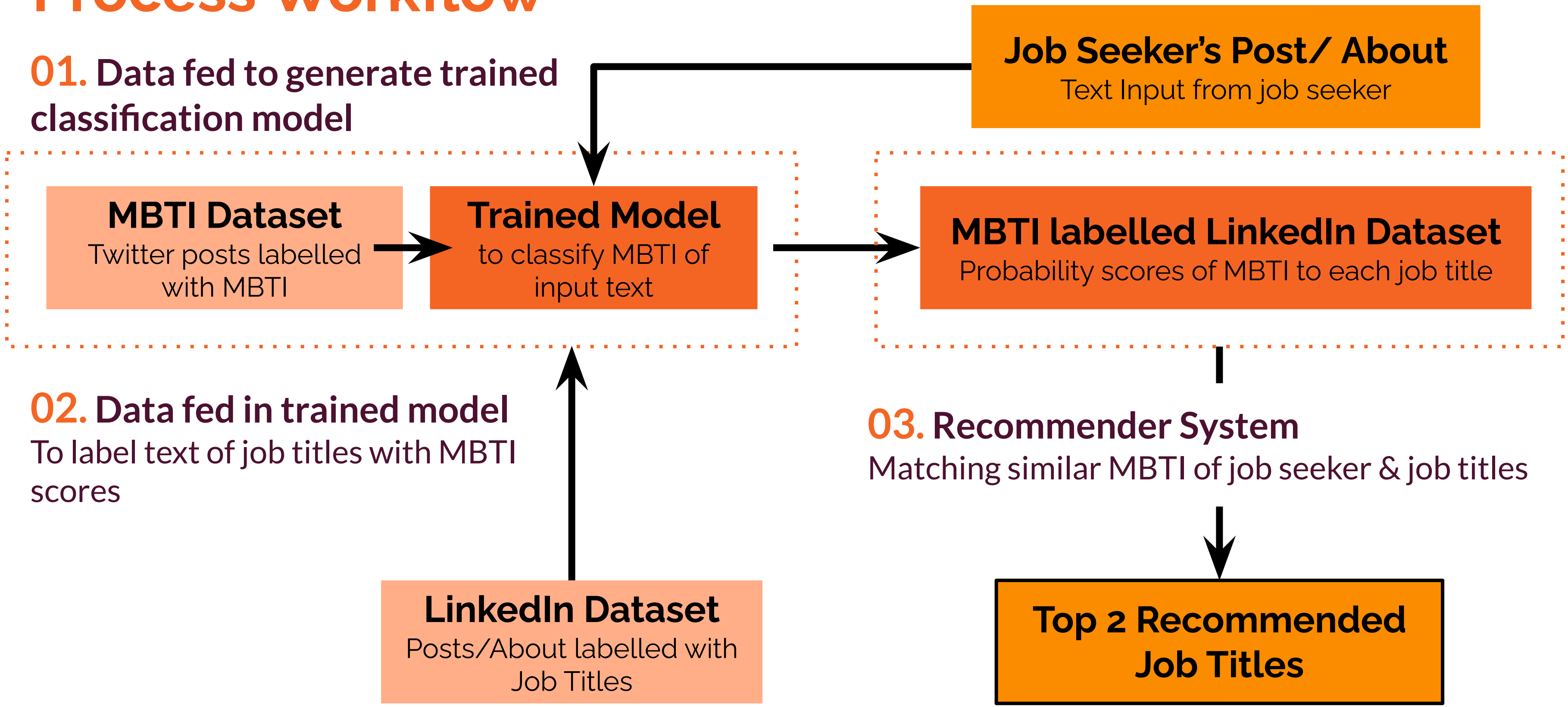
Problem Statement

Using MBTI personality traits, how could a mid-career switch job seeker like Sharon find a suitable role in the tech field, and remain satisfied in the new role?

Target Audience

- **Mid-career job seekers interested in tech industry**
  - Unsure of the exact roles to apply/ work towards
  - Do not have a clear picture of the non-tech skills (personality traits) demanded by the various tech roles
  - Needs to be employed quickly

Process Workflow



Data Collection/ Synthesis

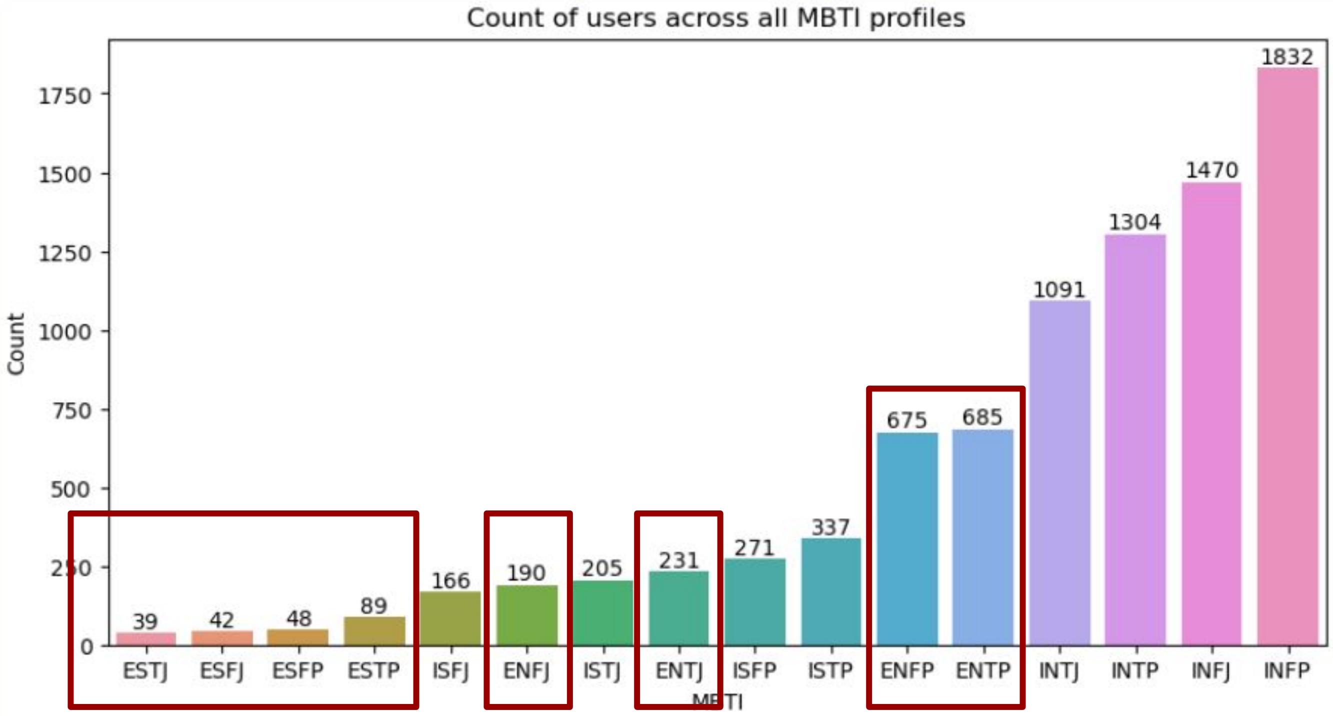
MBTI dataset

- Kaggle Dataset: Tweets labelled with MBTI
- Assessment of only introverted Personality traits

LinkedIn dataset

- Obtained raw data from looking up LinkedIn profiles and posts manually
- Creation of synthetic data to increase sample size

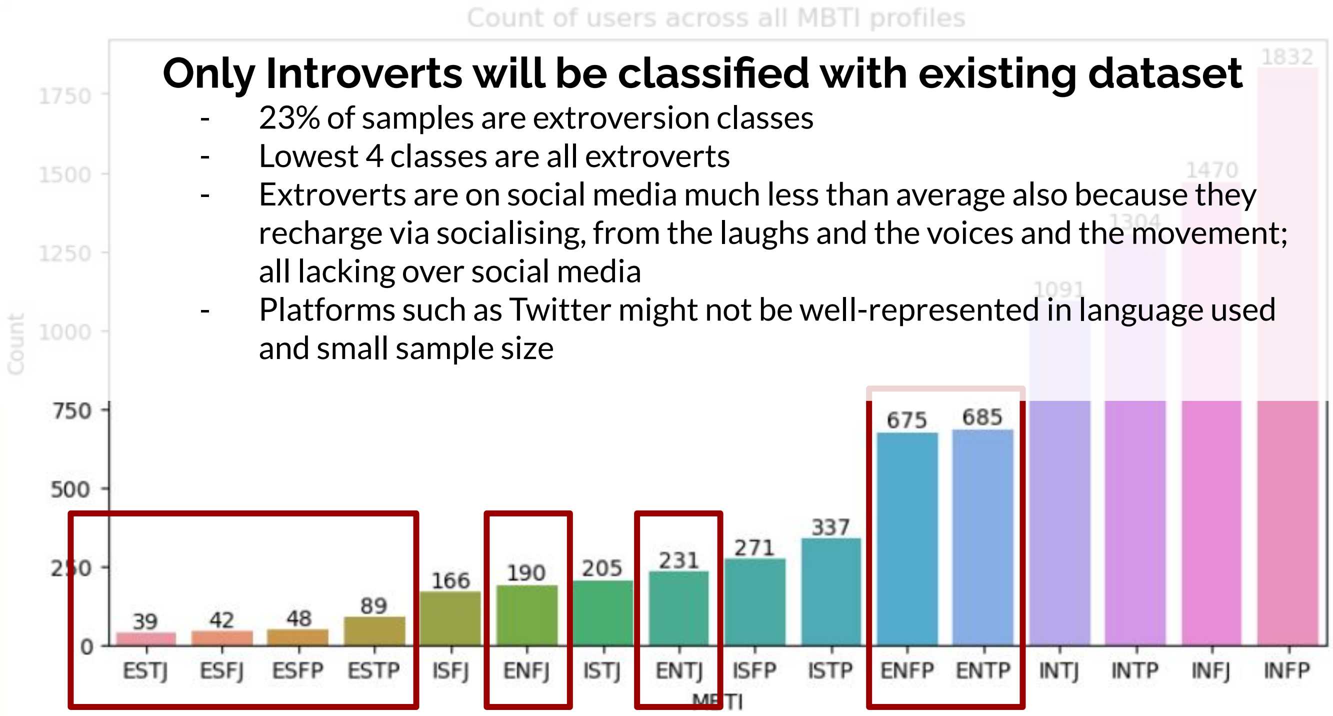
MBTI Dataset: Distribution of MBTI classes



MBTI Dataset: Distribution of MBTI classes

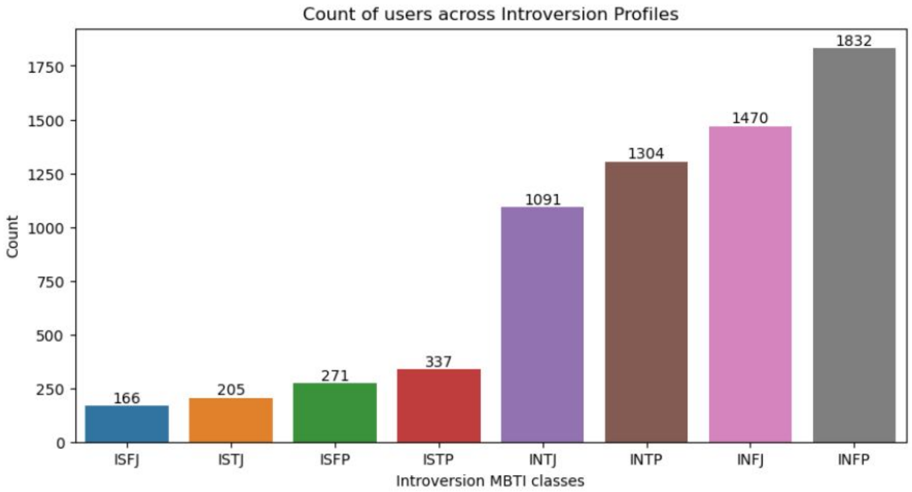
Only Introverts will be classified with existing dataset

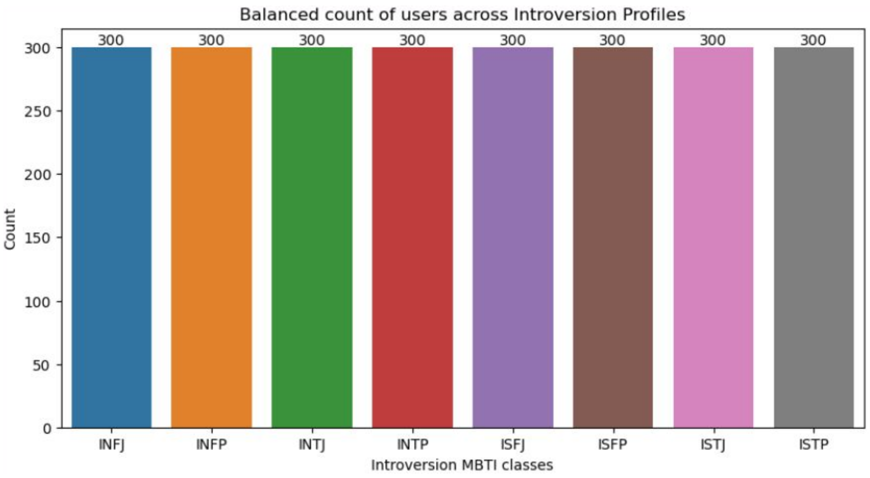
- 23% of samples are extroversion classes
- Lowest 4 classes are all extroverts
- Extroverts are on social media much less than average also because they recharge via socialising, from the laughs and the voices and the movement; all lacking over social media
- Platforms such as Twitter might not be well-represented in language used and small sample size



Before balancing

After balancing





- Mixture of random oversampling and undersampling conducted to achieve 300 tweets per introverted class

LinkedIn Dataset

1. **100 posts** - **Raw** LinkedIn Singaporean posts manually searched
    a. Limitation of stringent measures by LinkedIn to block any web scraping

2. 40 x **synthetic** posts each for 12 x job titles = **480 posts**
    a. **50 x generic 'drivers'** for a job titles permuted with **20 unique 'motivations'** specific to each job title

LinkedIn Dataset

Generic Drivers (x 50)
"the challenge of new problems"
"the satisfaction of client success"
"the continuous learning process",
…

Machine Learning Engineer: (x 20)

- "Developing cutting-edge machine learning models to solve industry challenges is my passion."
- "I specialize in applying deep learning techniques to enhance predictive modeling and analysis."
- …

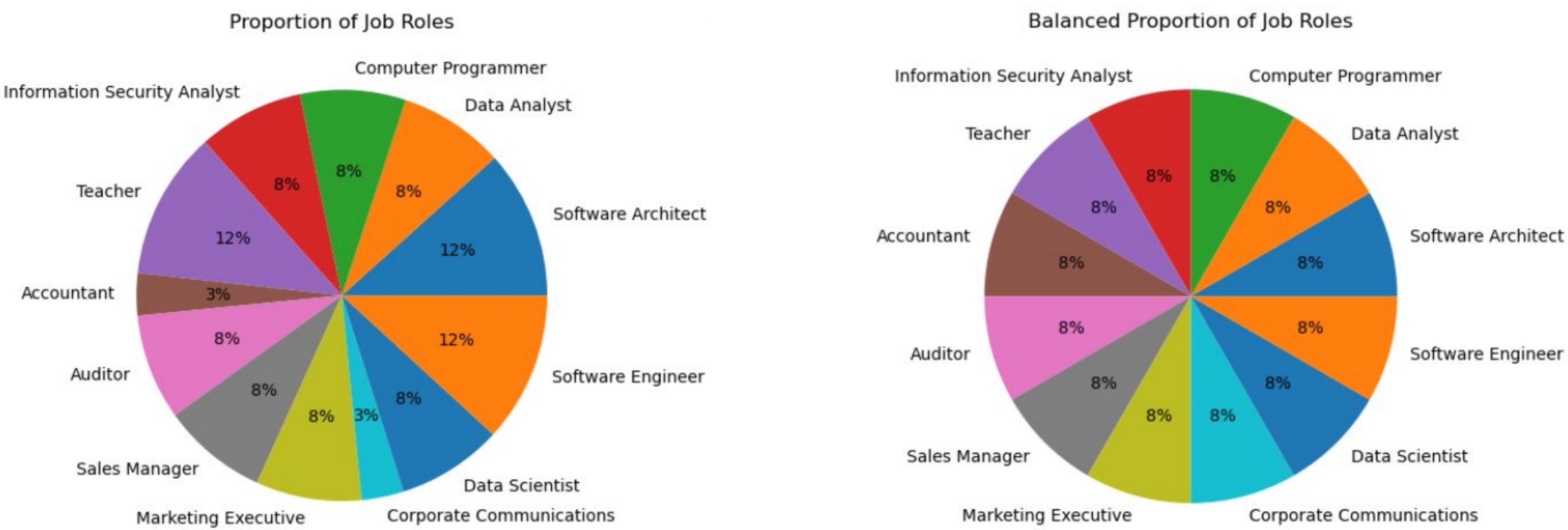
Computer Programmer: (x 20)

- "Crafting efficient, readable code that solves complex problems is my specialty."
- "I thrive on turning software designs into functional programming.",
- …

Final Posts: 20 posts each for 12 job titles

Machine Learning Engineer
"the challenge of new problems, developing cutting-edge machine learning models to solve industry challenges is my passion.”
…
Computer Programmer
"the challenge of new problems, I thrive on turning software designs into functional programming. "
…

LinkedIn Dataset: Distribution of Job Roles



- Mixture of random oversampling and undersampling conducted to achieve 20 posts tweets per job title

Data processing & cleaning

- Removed URL from posts
- Removed MBTI labels and Enneagram type labels from posts
- Remove non-english posts
- Remove unique characters, whitespaces, emojis, emojis hexadecimal codes

MBTI identification model

EDA - Cleaning (an example)

- Remove URLS [ ]
- Remove unique characters such as ||| or ___ [ ]
- Used regex

Initial	Good one _____ https://www.youtube.com/watch?v=fHiGbolFFGw|||Of course, to which I say I know
After	Good one Of course, to which I say I know

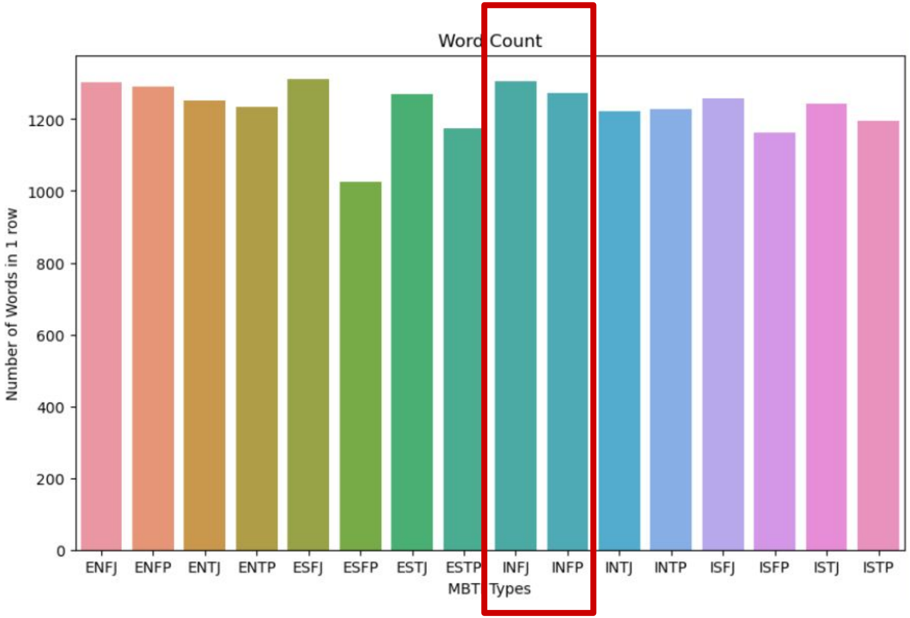
Other standardizations that we did:

- Standardising all apostrophes
- Remove extra white spaces
- Remove apostrophes in start and end of string
- Check for empty strings

MBTI identification model

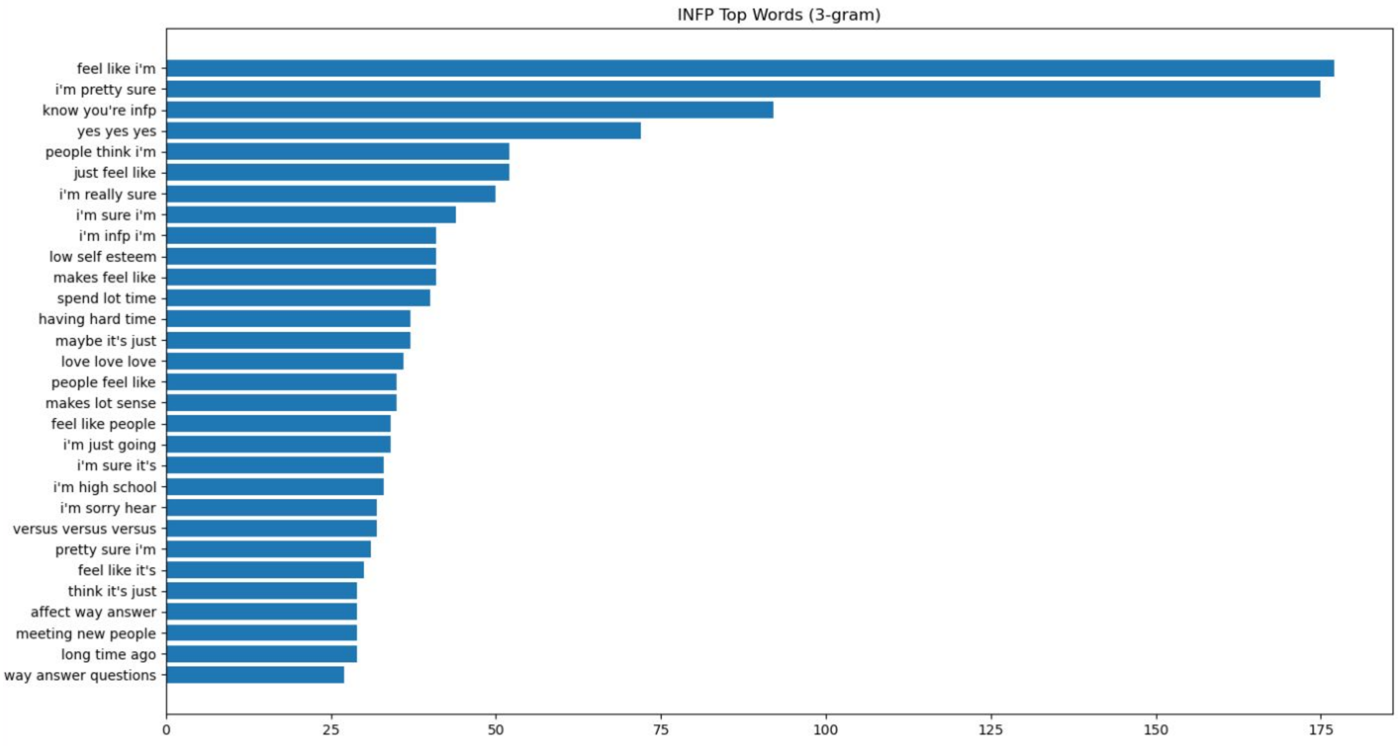
Other insights

- Previous chart we showed number of posts based on MBTI type - INFPs and INFJs were the highest
-
- Based on word count, all classes are not too different
- Frequency of posts makes a difference more than words typed



MBTI identification model

Other insights - INFP



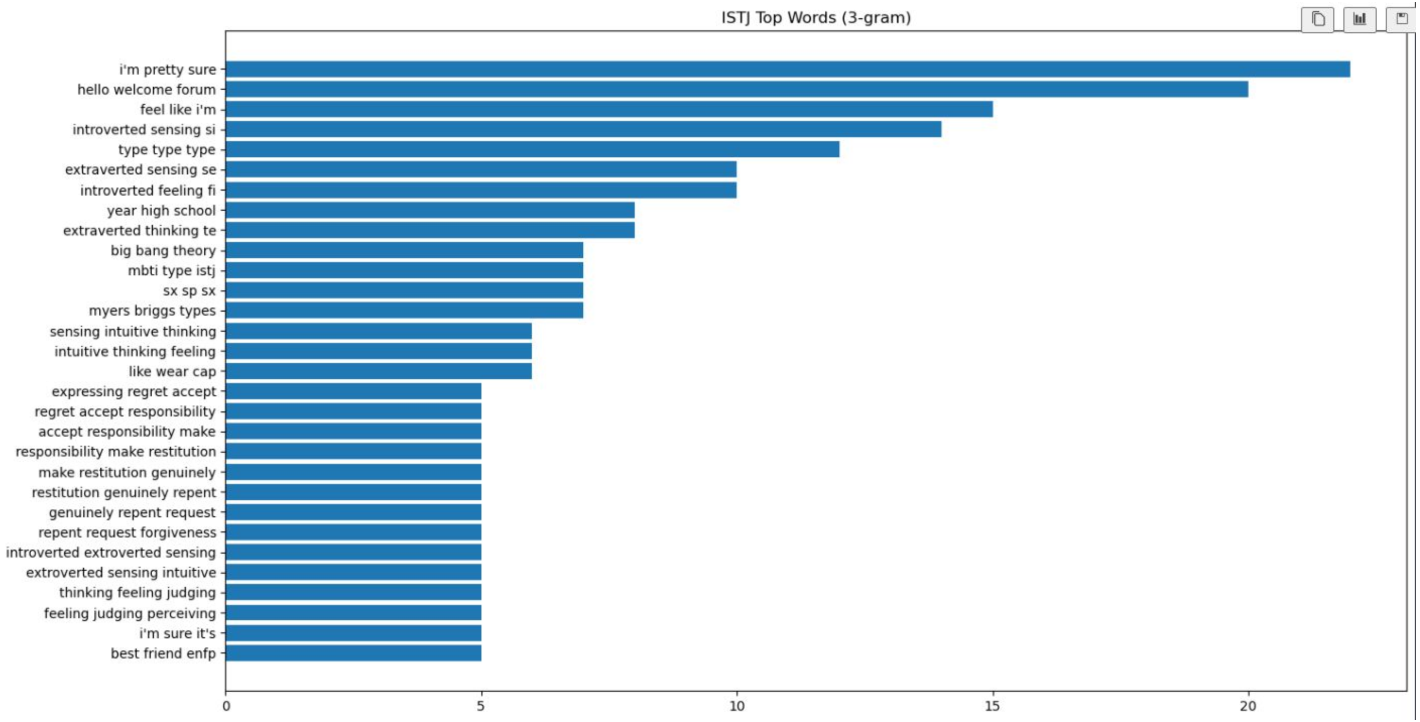
- Highest number of posts from this MBTI type

- More subjective words expressed

MBTI identification model

Other insights - ISTJ

- Lowest number of posts from this MBTI type
- More objective words expressed



MBTI identification model

Tokenizing & Stemming the words

Step 1: Remove apostrophe [ ]

Step 2: Tokenize the individual words [ ]

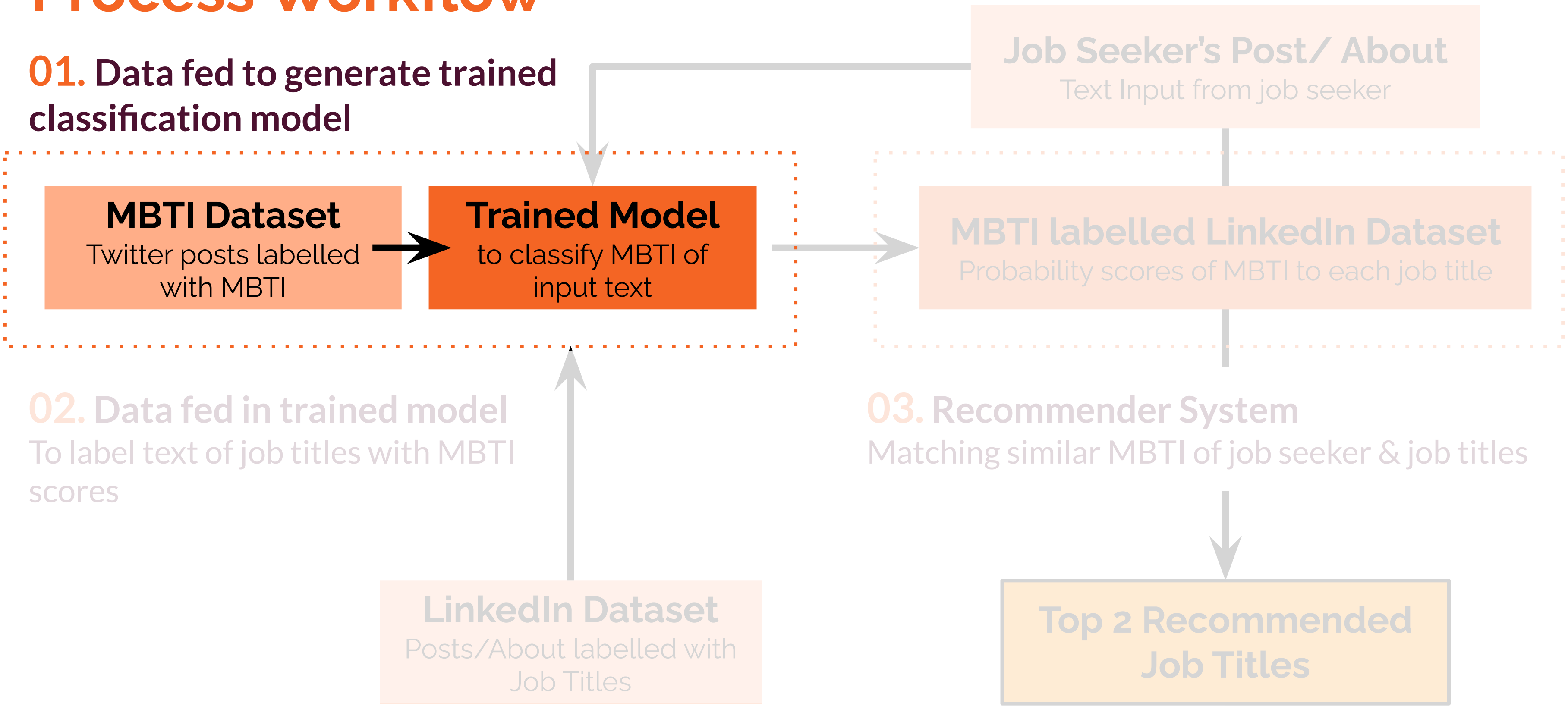
Step 3: Stem the individual words [ ]

Initial	i tend to build up a collection of things on my desktop that i use frequently and then move them into a folder called 'everything' from there it get sorted
After Step 1	i tend to build up a collection of things on my desktop that i use frequently and then move them into a folder called everything from there it get sorted
After Step 2	i tend to build up a collection of things on my desktop that i use frequently and then move them into a folder called everything from there it get sorted
After Step 3	i tend to build up collect of thing on my desktop that i use frequent and then move them into folder call everyth from there it get sort

Modeling Algorithms

- Logistic Regression
- Bernoulli Naive Bayes
- Multinomial Naive Bayes
- Support Vector Classifier (SVC)
- FNN

Process Workflow



Implemented Models

Model	Description
Logistic Regression	- Using probabilities to classify the predicted binary response with reference to a threshold probability e.g 0.5 mid-point - Relatively faster run time
Bernoulli Naive Bayes	- Classification algorithm relying on Bayes Theorem for binary data
Multinomial Naive Bayes	- Classification algorithm relying on Bayes Theorem for count data
SVC	- Classification algorithm for binary data under the umbrella of SVM - Finds an optimal hyperplane in N-dimension (N = no. of features) to distinctly classify the data points by looking for the maximum margin between opposing classes
Neural Network	- Unsupervised learning modelling algorithm where information flows in ONE-DIRECTION - Ultimately reach the final node where the prediction is made

Selected Metric: Accuracy

Accuracy:

$$\frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

Precision:

$$\frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

Recall:

$$\frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

F1-Score:

$$\frac{2 \text{ x (Precision x Recall)}}{\text{Precision + Recall}}$$

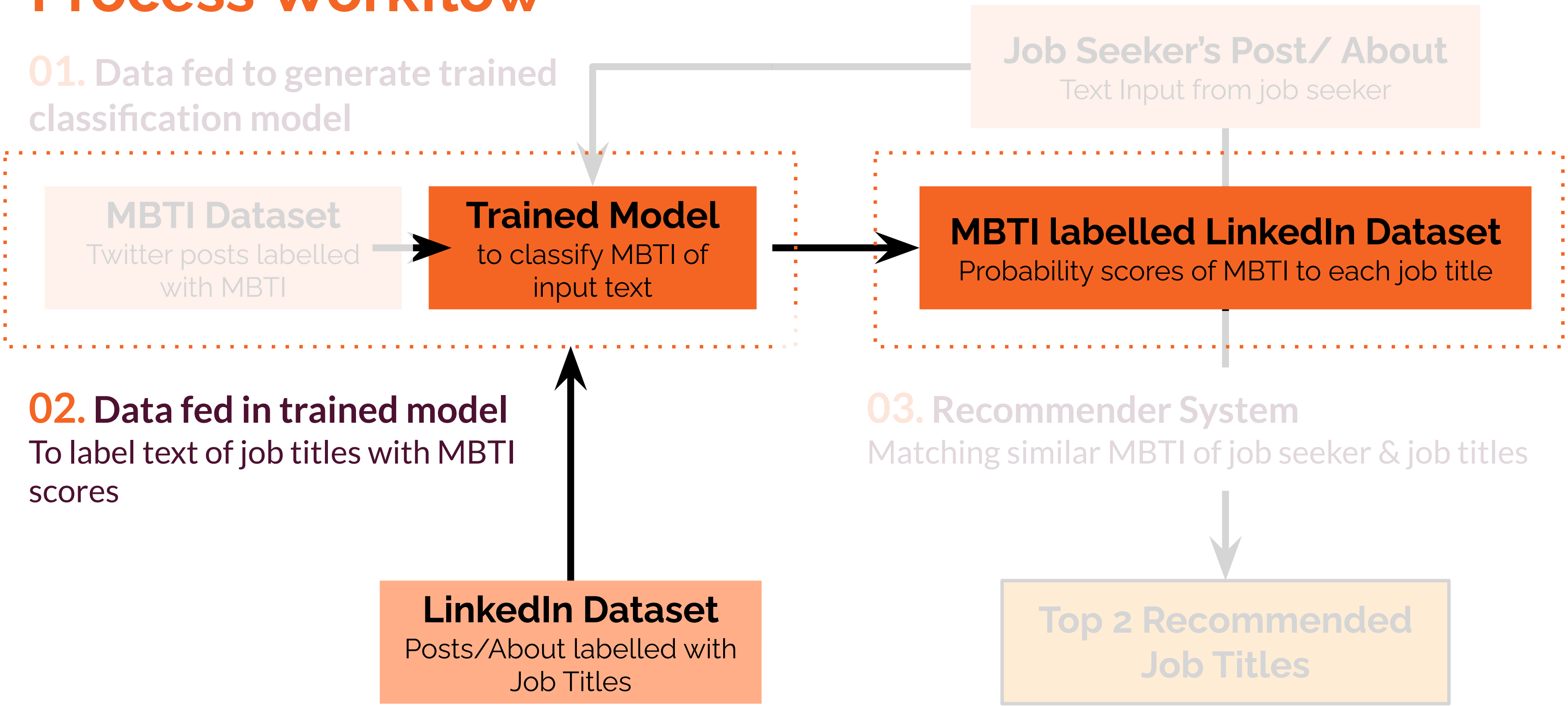
Rationale on score metric:

- Our goal is to accurately classify the MBTI of the user who posted a tweet - equal emphasis on all classes in each personality trait
- Other metrics such as recall/precision are not prioritised
- Selected data set is also relatively balanced

Results

Model Algorithm	Train Accuracy	Test Accuracy	Test F1	Train Accuracy	Test Accuracy	Test F1	Train Accuracy	Test Accuracy	Test F1
	S-N Trait			T-F Trait			J-P Trait		
Logistic Regression	0.939	0.883	0.882	0.931	0.903	0.904	0.917	0.847	0.848
Bernoulli Naive Bayes	0.786	0.720	0.720	0.798	0.747	0.741	0.734	0.668	0.666
Multinomial Naive Bayes	0.792	0.852	0.852	0.866	0.842	0.842	0.836	0.793	0.793
SVC	0.992	0.913	0.913	0.990	0.917	0.917	0.985	0.865	0.864
	Train Accuracy	Test Accuracy	Test Loss	Train Accuracy	Test Accuracy	Test Loss	Train Accuracy	Test Accuracy	Test Loss
NN	Up to 0.99	0.831	0.450	Up to 0.99	0.871	0.348	Up to 0.99	0.828	0.616*

Process Workflow

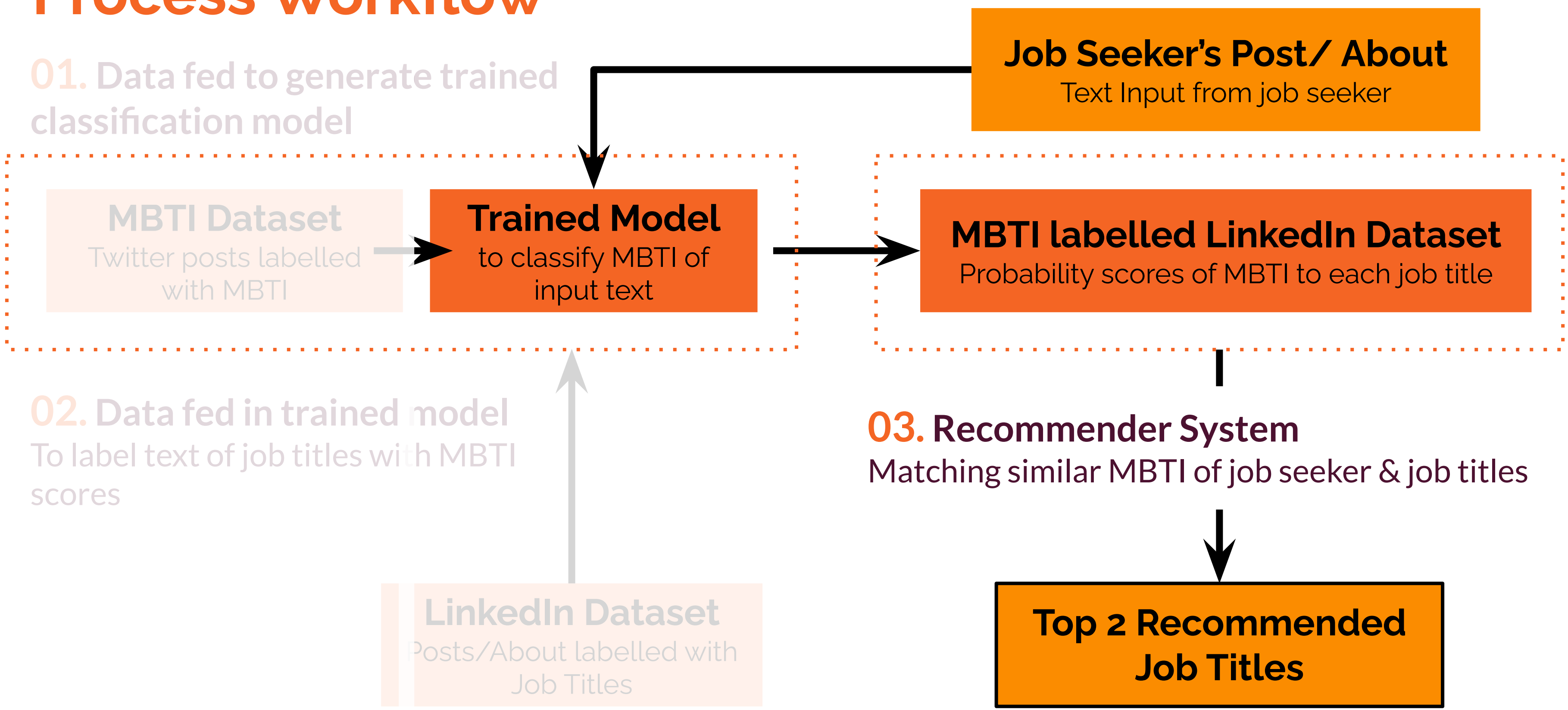


MBTI Classification Model to Label Job Titles with MBTI

Job Title	LinkedIn Post	S/N	T/F	J/P
Data Scientist	As a self-directed learner with my personal belief of leaving no regrets in life…	0.689	0.933	0.812
Software Engineer	Software Engineer interested in learning new tech stacks.	0.734	0.126	0.259
Software Engineer	Coding, for me, is more than a skill. it's a way of life. It's about distilling complex challenges into elegant solutions…	0.194	0.765	0.891
Data Scientist	Exploring the Evolution of Data Science: From Australopithecus to Homo Superior…	0.561	0.723	0.401
Accountant	A deep thinker who is goal-oriented and timeline driven, personally convicted to strive for improvement and perfection…	0.109	0.497	0.739

S - 1, N - 0; T - 1, F - 0; J - 1, P - 0

Process Workflow



Recommender System

Classify Sharon's MBTI probability values

S - 1, N - 0; T - 1, F - 0; J - 1, P - 0

Trained Model
to classify MBTI of input text

Sharon's Post	**S/N**	**T/F**	**J/P**
Curious about the impact of small, daily habits on propelling a data analyst's career forward? I've extracted compelling insights from 'Atomic Habits' by James Clear, tailored specifically for those in the data field.	0.34	0.31	0.47

Recommender System

Identify job titles with closest MBTI scores with Sharon's, using euclidean distance as a measure - smaller distances indicate higher similarities

MBTI labelled LinkedIn Dataset

Probability scores of MBTI to each job title

Sharon's MBTI		
S/N	**T/F**	**J/P**
0.34	0.31	0.47

Job Title	**S/N**	**T/F**	**J/P**
Teacher	0.68	0.93	0.81
Data Scientist	0.30	0.21	0.6
...	...	...	...

Euclidean Distance
0.784
0.169
...

Recommender System

Distances will be arranged in ascending order, taking only the shortest 20 distances and job titles

Job Title	Euclidean Distance
Data Scientist	0.169
Data Scientist	0.187
Data Scientist	0.191
Data Analyst	0.312
Data Analyst	0.351
...	...

Job Title	Average Euclidean Distance
Data Scientist	(0.169 + 0.187 + 0.191) / 3 = 0.182
Data Analyst	(0.312 + 0.351) / 2 = 0.331
...	...

For each job title category, distances are aggregated and normalised with job titles' frequency of appearance in the 20 shortlisted titles. 2 titles with the smallest final distances will be recommended.

Cost Benefit Analysis: Cost

1. Cost of Burnout per Employee
- Direct Costs
    - According to our research, the cost of burnout could be **as much as half** an employee's annual salary*
    - For an employee earning $50,000, this would be $25,000

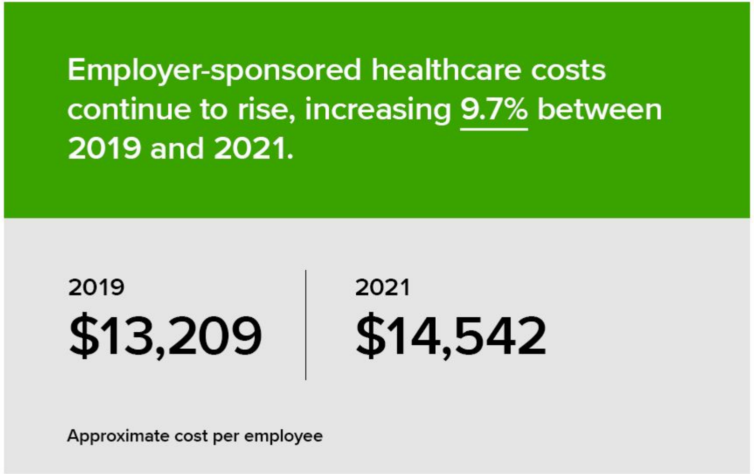
2. Productivity Loss
    - Disengaged employees, often a result of burnout, cost their employer an **average of 34% of their annual salary** due to lost productivity*
    - For a $50,000 salary, this amounts to $17,000.
    - Burnout resulted in a productivity **loss of 4.2 hours** per week in Singapore

*References at end of slides

Cost Benefit Analysis: Cost

- Healthcare Costs
    - Burnout can lead to increased healthcare costs
    - More spending on employees with chronic conditions, which can be exacerbated by burnout*
    - In Singapore, cost is estimated to be around **US$2.3 billion***

The Costly Consequences of Employee Burnout



Cost Benefit Analysis: Cost

2. Total Cost of Burnout for the Company

- Lost Productivity
    - The U.S. lost $1.8 trillion in productivity due to corporate burnout
    - **38% of resignations in the tech industry*** in Singapore are attributed to stress and burnout

FORBES > MONEY

Corporate Burnout Is Coming For Investor Profits

Q.ai - Powering a Personal Wealth Movement Former Contributor

Making wealth creation easy, accessible and transparent.

*References at end of slides

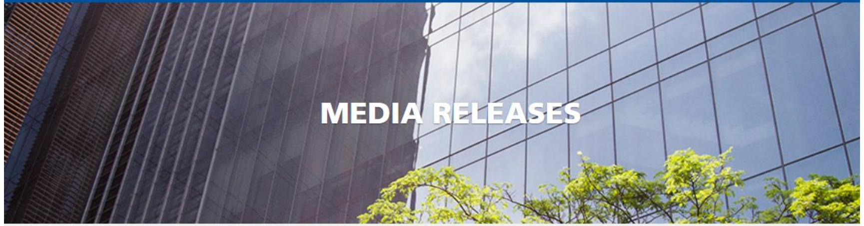
Cost Benefit Analysis: Cost

- Turnover Costs
  - Replacing an employee can cost from **1.5 to 2x** the employee's salary
  - For a tech company with a high incidence of burnout, these costs can be substantial
  - In Singapore, it is estimated to cost Singapore **almost $12 billion***

DukeNUS Medical School

COVID19 STORIES APPLY GIVING DIRECTORY ALUMNI LOGIN

ABOUT ADMISSIONS EDUCATION RESEARCH ACADEMIC MEDICINE INNOVATION



Duke-NUS Medical School > News > Duke-NUS, IMH: Cost of anxiety and depression in Singapore runs into the billions

BACK

Wednesday, 26 Apr, 2023

Duke-NUS, IMH: Cost of anxiety and depression in Singapore runs into the billions

Singapore scientists find use of healthcare resources, absenteeism and reduced productivity due to mental health issues place a significant burden on Singapore's economy.

SINGAPORE, 26 April 2023 – Symptoms of anxiety and depression in the post-peak pandemic era could be costing Singapore 2.9 per cent of its gross domestic product (GDP)—or nearly S$16 billion—suggests a study conducted by Duke-NUS Medical School and the Institute of Mental Health

EMPLOYEES WITH A HIGH RISK OF DEVELOPING MENTAL HEALTH ISSUES

LOCATION	PERCENTAGE
SOUTH KOREA	44%
MALAYSIA	42%
JAPAN	41%
HONG KONG	40%
CHINA	39%
INDIA	39%
VIETNAM	35%
TAIWAN	35%
SINGAPORE	34%

Cost Benefit Analysis: Benefits

1. Savings from Reduced Turnover

- Effective job fit assessments can reduce turnover-related expenses significantly
- Crucial
    - turnover costs can equal up to one-third of the employee's annual salary
    - job fit assessments **reduce turnover by 29% to 59%***

2. Value from Increased Productivity

- Productivity Increase:
    - Engaged employees are more productive
    - **Conservative 20% increase** in productivity for an employee earning $50,000 would equate to an additional $10,000 in value per employee.*
    - In Singapore, the actual cost and savings will vary depending on effectiveness of measures to curb burnout

While the estimates are based on available data and reports, actual costs and savings will vary depending on specific company circumstances and the effectiveness of implemented measures to combat burnout and improve job fit*

*References at end of slides

Compounding the Effects

Improved job satisfaction of job seekers

- Mid-career job seekers are more quickly employed to continue supporting their families
- Resources (both time and money) saved from taking unnecessary upskill courses and training
- Improved job satisfaction and allow greater meaning in work for job seeker

Minimising unemployment burden

- With suitable job roles quickly identified, unnecessary mismatched training could be avoided, reducing subsidies spent by WoG
- Improve in job satisfaction allows improve in mental health of seekers in further improving productivity of economy

Compounding the Effects

Increasing productivity in companies

- Reduction in downtime of ‘handholding’ new staff
- Enhancing innovation as mid-career professionals combine skills/ fresh perspectives from past vocations to present roles
- Higher retention rate - avoids inefficient use of resources to re-train new staff constantly
- Mid-career seekers could be more adaptive to change, a critical trait in tech industry
- Company’s productivity and performance increases

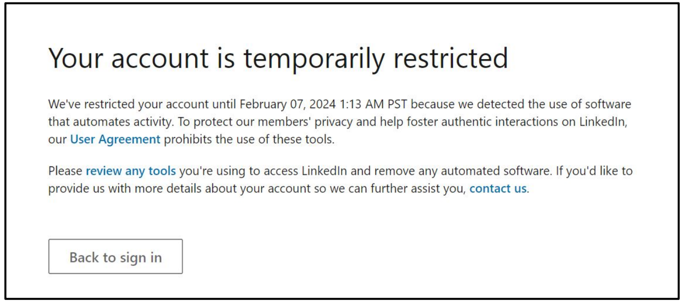
#Lifelong_learning

- Training plans tailored to employee’s personality (help them learn in the method that suits them best)
- Ensuring company is always ahead to the most relevant skills for best performance (Skill learning becomes diversified)

Limitations in Data Collection

- LinkedIn Dataset consists of synthetic data as LinkedIn is highly stringent and against web scraping their user's profiles
    - MBTI and job titles in Recommender System might not be representative of the actual MBTI population/ proportion in job titles

- Training MBTI Classification model using tweets might be limited in labelling other text forms e.g. LinkedIn posts as language used could be different



Future Work

- Tap on existing employees' social media posts
  - to further improve classification of MBTI types and recommender system
- Collect more data for extroverts and retrain the model to include prediction for both introverts and extroverts
- Explore model training on other personality tests
  i.e. Enneagram, DISC personality tests to see if there are similar effects
  - Useful for companies which do not use MBTI but other form of personality tests
- Explore models such as CNN/ RNN in improving MBTI classification model

Acknowledgements

- Our instructor and TA for listening to our ideas & giving feedback & troubleshooting our codes with us

- Our classmates who
  - who graciously shared their text cleaning codes from their previous projects (#slay)

  - explained what is dot product to us and bouncing ideas off together

  - Structuring previous notebooks so nicely that we could adopt the same format

- #bestie , google, stackoverflow, TDS articles & more

- All the coffee we bought and drank from different stalls in Tanjong Pagar

References

- https://www.edume.com/blog/cost-of-training-a-new-employee
- https://www.auston.edu.sg/advice/subsidies-for-skillsfuture-courses-in-singapore/
- https://www.skillsfuture.gov.sg/initiatives/mid-career/enhancedsubsidy#:~:text=What%20is%20it%3F,adapt%20to%20changing%20job%20requirements.
- https://hrshelf.com/cost-of-training-employees/
- https://vervoe.com/the-real-cost-of-training-a-new-hire/
- https://elmlearning.com/blog/how-much-does-employee-training-really-cost/
- https://research.aimultiple.com/linkedin-scrapers/
- https://www.phdata.io/blog/what-is-the-cost-to-deploy-and-maintain-a-machine-learning-model/
- https://medium.com/@yomna/im-99-extroverted-and-i-struggle-with-social-media-f627c40982f1
- https://www.3treetech.com/cost-of-burnout-in-us-companies-in-2023-burnout-cost/
- https://www.whoopunite.com/blog/business/articles/cost-of-employee-burnout/
- https://www.forbes.com/sites/qai/2023/01/30/corporate-burnout-is-coming-for-investor-profits/?sh=602d322f7008
- https://builtin.com/recruiting/cost-of-turnover
- https://theindependent.sg/8-in-10-it-professionals-experience-burnout-survey/
- https://www.zdnet.com/google-amp/article/cybersecurity-burnout-hits-apac-firms-with-lack-of-resources-the-key-challenge/
- https://www.cigna.com.sg/health-content-hub/mental-health/how-work-related-stress-affects-employee
- https://www.cnbc.com/2023/09/20/4-in-5-employees-in-asia-at-risk-developing-mental-health-issues-study.html
- https://www.graebel.com/blog/graebels-innovative-solution-lowers-costs-enhances-employee-productivity-in-singapore/