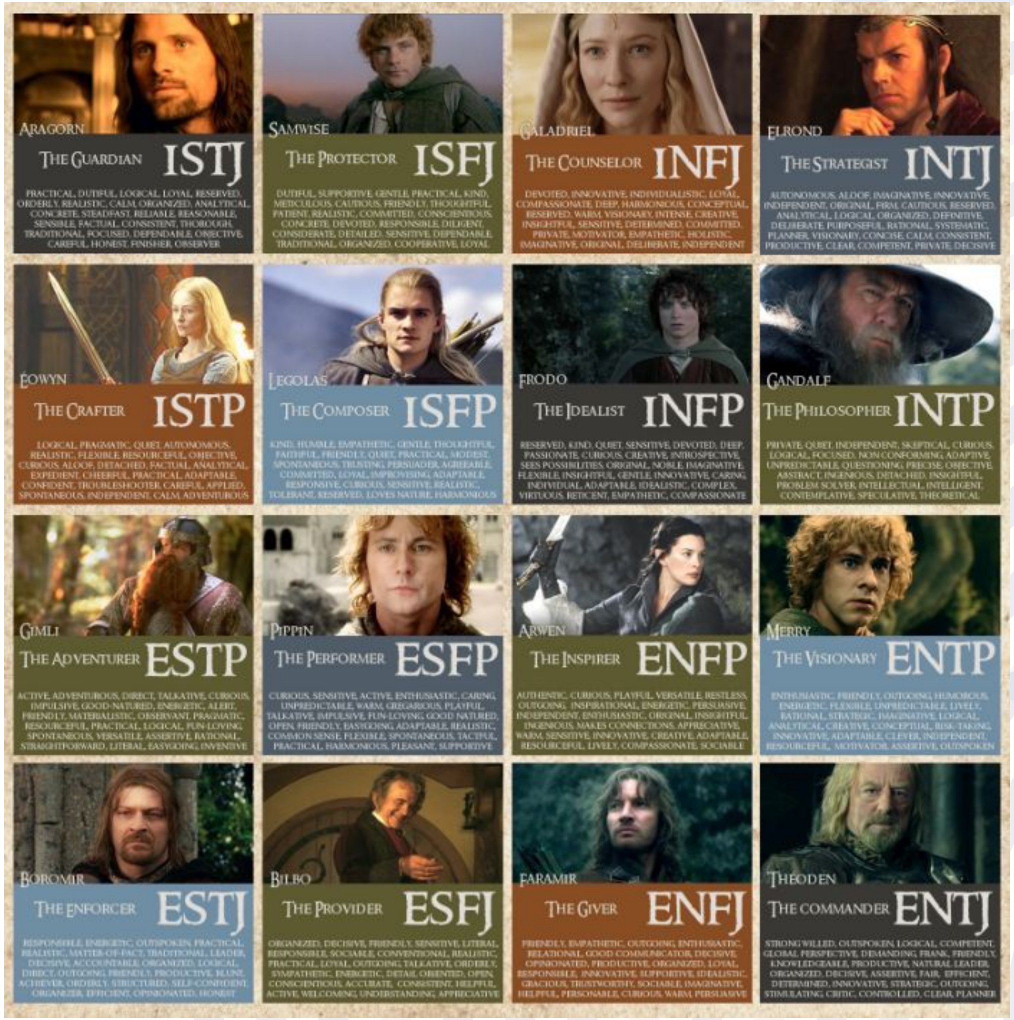
MBTI Type Finder

ASDML SS21

Contents

The Myers-Briggs Type Indicator



ARAGORN
THE GUARDIAN ISTJ
PRACTICAL, DUTIFUL, LOGICAL, LOYAL, RESERVED, ORDERLY, REALISTIC, CALM, ORGANIZED, ANALYTICAL, CONCRETE, STEADFAST, RELIABLE, REASONABLE, SENSIBLE, FACTUAL, CONSISTENT, THOROUGH, TRADITIONAL, FOCUSED, DEPENDABLE, OBJECTIVE, CAREFUL, HONEST, FINISHER, OBSERVER
SAMWISE
THE PROTECTOR ISFJ
DUTIFUL, SUPPORTIVE, GENTLE, PRACTICAL, KIND, METICULOUS, CAUTIOUS, FRIENDLY, THOUGHTFUL, PATIENT, REALISTIC, COMMITTED, CONSCIENTIOUS, CONCRETE, DEVOTED, RESPONSIBLE, DILIGENT, CONSIDERATE, DETAILED, SENSITIVE, DEPENDABLE, TRADITIONAL, ORGANIZED, COOPERATIVE, LOYAL
GALADRIEL
THE COUNSELOR INFJ
DEVOTED, INNOVATIVE, INDIVIDUALISTIC, LOYAL, COMPASSIONATE, DEEP, HARMONIOUS, CONCEPTUAL, RESERVED, WARM, VISIONARY, INTENSE, CREATIVE, INSIGHTFUL, SENSITIVE, DETERMINED, COMMITTED, PRIVATE, MOTIVATOR, EMPATHETIC, HOLISTIC, IMAGINATIVE, ORIGINAL, DELIBERATE, INDEPENDENT
ELROND
THE STRATEGIST INTJ
AUTONOMOUS, ALOOF, IMAGINATIVE, INNOVATIVE, INDEPENDENT, ORIGINAL, FIRM, CAUTIOUS, RESERVED, ANALYTICAL, LOGICAL, ORGANIZED, DEFINITIVE, DELIBERATE, PURPOSEFUL, RATIONAL, SYSTEMATIC, PLANNER, VISIONARY, CONCISE, CALM, CONSISTENT, PRODUCTIVE, CLEAR, COMPETENT, PRIVATE, DECISIVE
ÉOWYN
THE CRAFTER ISTP
LOGICAL, PRAGMATIC, QUIET, AUTONOMOUS, REALISTIC, FLEXIBLE, RESOURCEFUL, OBJECTIVE, CURIOUS, ALOOF, DETACHED, FACTUAL, ANALYTICAL, EXPEDIENT, CHEERFUL, PRACTICAL, ADAPTABLE, CONFIDENT, TROUBLESHOOTER, CAREFUL, APPLIED, SPONTANEOUS, INDEPENDENT, CALM, ADVENTUROUS
LEGOLAS
THE COMPOSER ISFP
KIND, HUMBLE, EMPATHETIC, GENTLE, THOUGHTFUL, FAITHFUL, FRIENDLY, QUIET, PRACTICAL, MODEST, SPONTANEOUS, TRUSTING, PERSUADER, AGREEABLE, COMMITTED, LOYAL, IMPROVISING, ADAPTABLE, RESPONSIVE, CURIOUS, SENSITIVE, REALISTIC, TOLERANT, RESERVED, LOVES NATURE, HARMONIOUS
FRODO
THE IDEALIST INFP
RESERVED, KIND, QUIET, SENSITIVE, DEVOTED, DEEP, PASSIONATE, CURIOUS, CREATIVE, INTROSPECTIVE, SEES POSSIBILITIES, ORIGINAL, NOBLE, IMAGINATIVE, FLEXIBLE, INSIGHTFUL, GENTLE, INNOVATIVE, CARING, INDIVIDUAL, ADAPTABLE, IDEALISTIC, COMPLEX, VIRTUOUS, RETICENT, EMPATHETIC, COMPASSIONATE
GANDALF
THE PHILOSOPHER INTP
PRIVATE, QUIET, INDEPENDENT, SKEPTICAL, CURIOUS, LOGICAL, FOCUSED, NON CONFORMING, ADAPTIVE, UNPREDICTABLE, QUESTIONING, PRECISE, OBJECTIVE, ABSTRACT, INGENIOUS, DETACHED, INSIGHTFUL, PROBLEM SOLVER, INTELLECTUAL, INTELLIGENT, CONTEMPLATIVE, SPECULATIVE, THEORETICAL
GIMLI
THE ADVENTURER ESTP
ACTIVE, ADVENTUROUS, DIRECT, TALKATIVE, CURIOUS, IMPULSIVE, GOOD-NATURED, ENERGETIC, ALERT, FRIENDLY, MATERIALISTIC, OBSERVANT, PRAGMATIC, RESOURCEFUL, PRACTICAL, LOGICAL, FUN-LOVING, SPONTANEOUS, VERSATILE, ASSERTIVE, RATIONAL, STRAIGHTFORWARD, LITERAL, EASYGOING, INVENTIVE
PIPPIN
THE PERFORMER ESFP
CURIOUS, SENSITIVE, ACTIVE, ENTHUSIASTIC, CARING, UNPREDICTABLE, WARM, GREGARIOUS, PLAYFUL, TALKATIVE, IMPULSIVE, FUN-LOVING, GOOD NATURED, OPEN, FRIENDLY, EASYGOING, ADAPTABLE, REALISTIC, COMMON SENSE, FLEXIBLE, SPONTANEOUS, TACTFUL, PRACTICAL, HARMONIOUS, PLEASANT, SUPPORTIVE
ARWEN
THE INSPIRER ENFP
AUTHENTIC, CURIOUS, PLAYFUL, VERSATILE, RESTLESS, OUTGOING, INSPIRATIONAL, ENERGETIC, PERSUASIVE, INDEPENDENT, ENTHUSIASTIC, ORIGINAL, INSIGHTFUL, INGENIOUS, MAKES CONNECTIONS, APPRECIATIVE, WARM, SENSITIVE, INNOVATIVE, CREATIVE, ADAPTABLE, RESOURCEFUL, LIVELY, COMPASSIONATE, SOCIABLE
MERRY
THE VISIONARY ENTP
ENTHUSIASTIC, FRIENDLY, OUTGOING, HUMOROUS, ENERGETIC, FLEXIBLE, UNPREDICTABLE, LIVELY, RATIONAL, STRATEGIC, IMAGINATIVE, LOGICAL, ANALYTICAL, CREATIVE, CONCEPTUAL, RISK-TAKING, INNOVATIVE, ADAPTABLE, CLEVER, INDEPENDENT, RESOURCEFUL, MOTIVATOR, ASSERTIVE, OUTSPOKEN
BOROMIR
THE ENFORCER ESTJ
RESPONSIBLE, ENERGETIC, OUTSPOKEN, PRACTICAL, REALISTIC, MATTER-OF-FACT, TRADITIONAL, LEADER, DECISIVE, ACCOUNTABLE, ORGANIZED, LOGICAL, DIRECT, OUTGOING, FRIENDLY, PRODUCTIVE, BLUNT, ACHIEVER, ORDERLY, STRUCTURED, SELF-CONFIDENT, ORGANIZER, EFFICIENT, OPINIONATED, HONEST
BILBO
THE PROVIDER ESFJ
ORGANIZED, DECISIVE, FRIENDLY, SENSITIVE, LITERAL, RESPONSIBLE, SOCIABLE, CONVENTIONAL, REALISTIC, PRACTICAL, LOYAL, OUTGOING, TALKATIVE, ORDERLY, SYMPATHETIC, ENERGETIC, DETAIL ORIENTED, OPEN, CONSCIENTIOUS, ACCURATE, CONSISTENT, HELPFUL, ACTIVE, WELCOMING, UNDERSTANDING, APPRECIATIVE
FARAMIR
THE GIVER ENFJ
FRIENDLY, EMPATHETIC, OUTGOING, ENTHUSIASTIC, RELATIONAL, GOOD COMMUNICATOR, DECISIVE, OPINIONATED, PRODUCTIVE, ORGANIZED, LOYAL, RESPONSIBLE, INNOVATIVE, SUPPORTIVE, IDEALISTIC, GRACIOUS, TRUSTWORTHY, SOCIABLE, IMAGINATIVE, HELPFUL, PERSONABLE, CURIOUS, WARM, PERSUASIVE
THÉODEN
THE COMMANDER ENTJ
STRONG WILLED, OUTSPOKEN, LOGICAL, COMPETENT, GLOBAL PERSPECTIVE, DEMANDING, FRANK, FRIENDLY, KNOWLEDGEABLE, PRODUCTIVE, NATURAL LEADER, ORGANIZED, DECISIVE, ASSERTIVE, FAIR, EFFICIENT, DETERMINED, INNOVATIVE, STRATEGIC, OUTGOING, STIMULATING, CRITIC, CONTROLLED, CLEAR, PLANNER

MBTI - Origin

- Personality test developed in 1962
- Used for HR purposes

Categorization

Extraversion (E):

Focus on outer world

Sensing (S):

Focus on basic information, on the presence

Thinking (T):

Look at logic and consistency

Judging (J):

Prefer to get things decided, plan in advance, organize

Introversion (I):

Focus on inner world

Intuition (N):

Prefer to interpret and add meaning, focus on the future

Feeling (F):

Look at people and circumstances

Perceiving (P):

stay flexible and open to new information, improvise

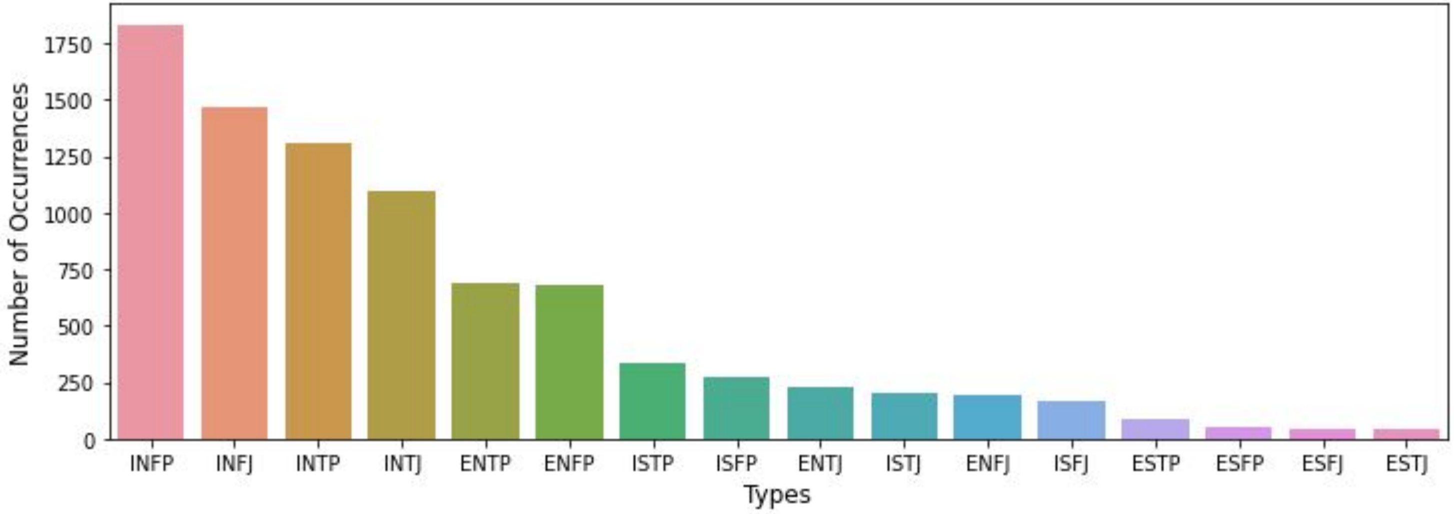
The Data Set

The Dataset

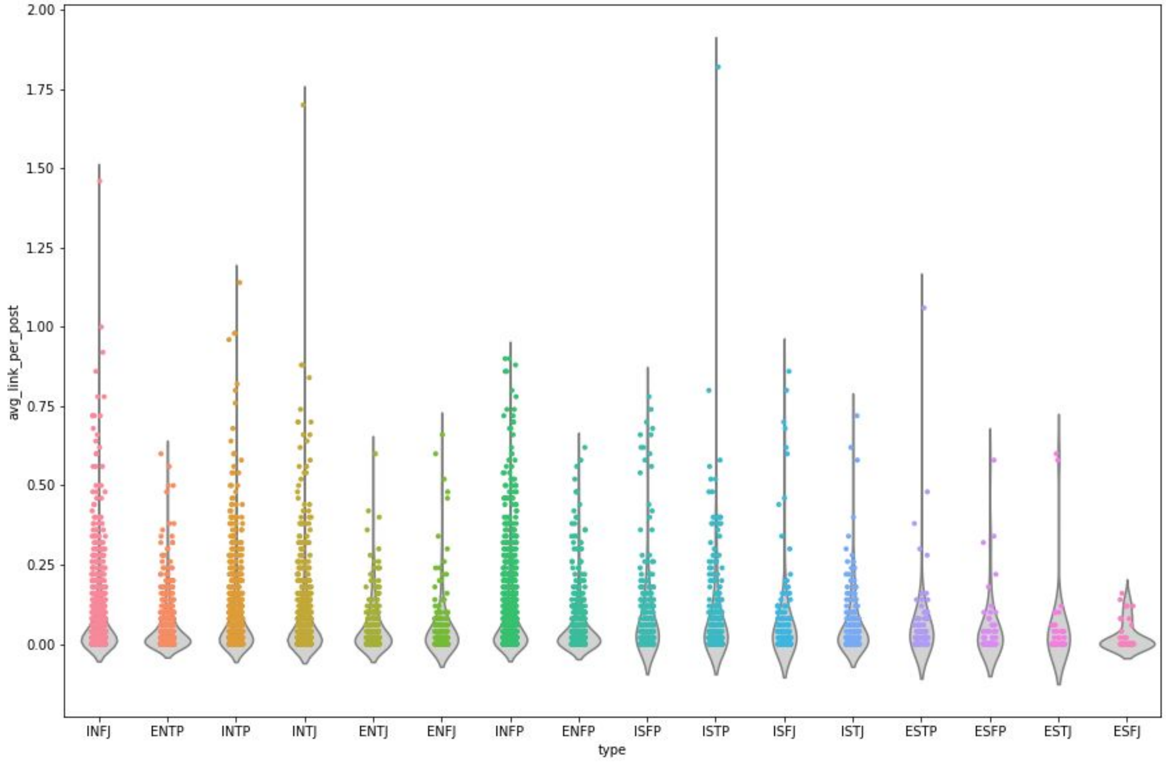
- Myers-Briggs Personality Type Dataset (kaggle)
- Collected through *personalitycafe.com*
- 8675 rows of data
  - Person's MBTI type
  - Last 50 posts

Number of posts of all MBTI types

- Unbalance in Introvert/Extrovert and INtuition/Sensing pairs

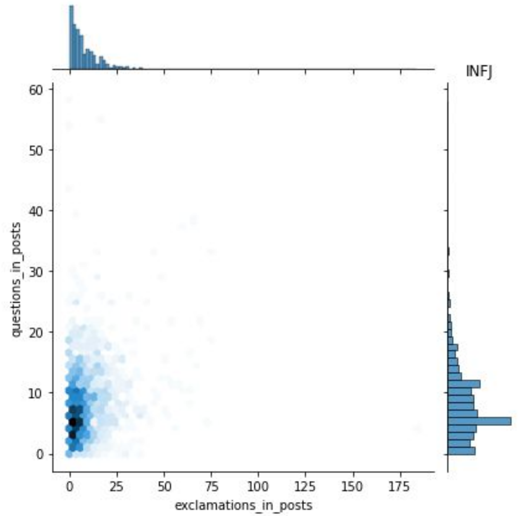


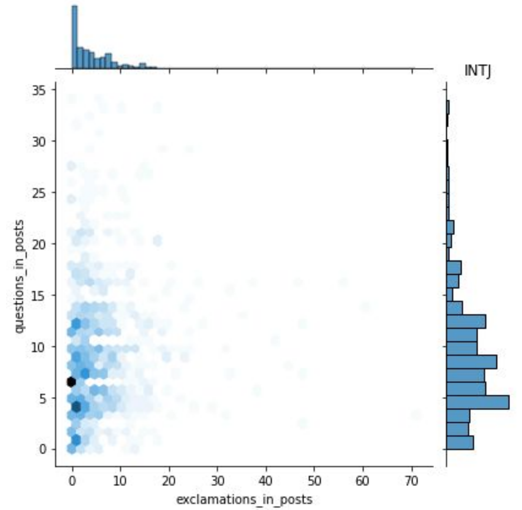
Average links per post



- More links posted by Introverts than Extroverts
- Other personality aspects relatively balanced

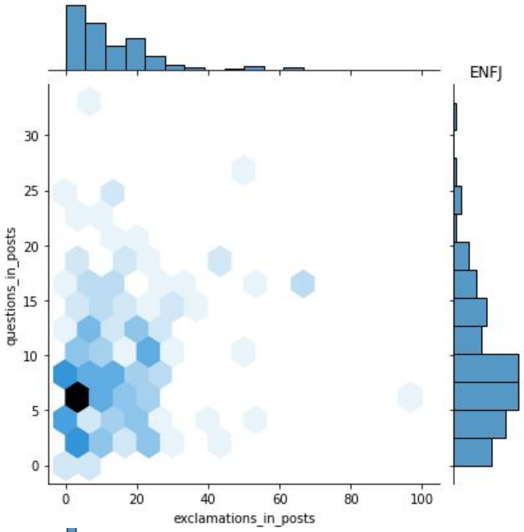
Questions vs. Exclamations

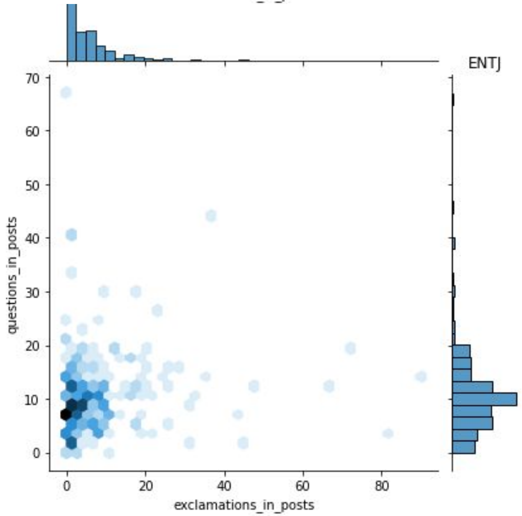




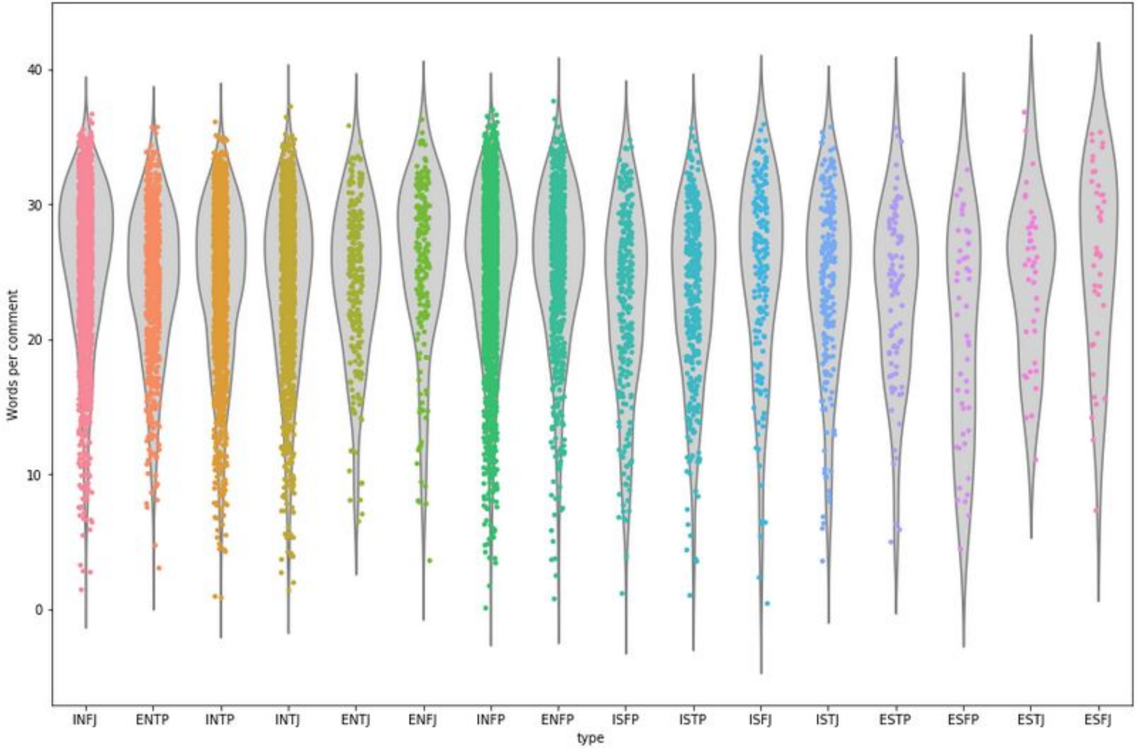
I vs. E

- Generally more question than exclamation marks
- Slightly more EMs for Extroverts





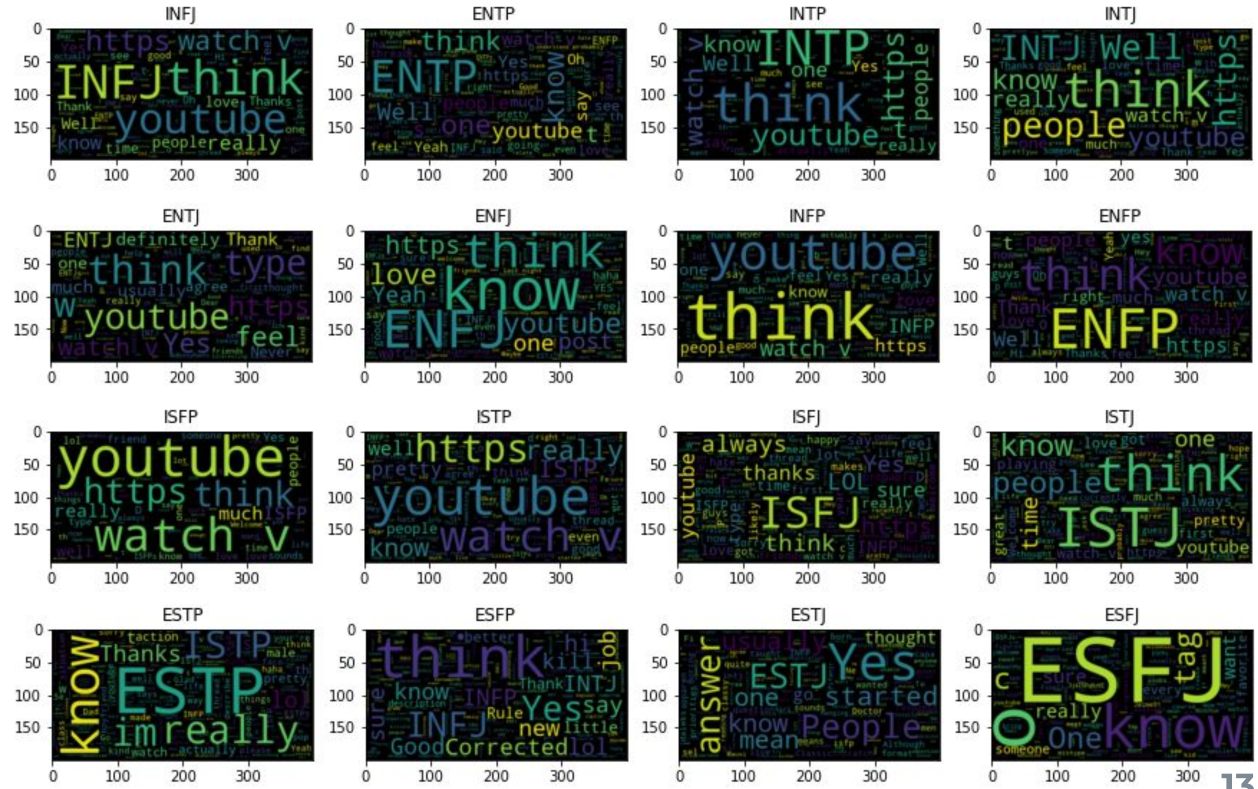
Average number of words per post



- F types have broader range between few and many words compared to T
- Otherwise relatively balanced

Word Clouds

- Many talk frequently about their own type (due to forum)
- Differences can be seen between types, but not really between personality aspects



Text Processing

Text Processing

- Counted and removed links
  - Improvement : replacing with <LINK>-keyword
- Expanded contractions (ain't" -> "are not, " s" -> " is)
- Counted and removed punctuations
  - Improvement : keeping them
- Removed numbers and words with numbers ("gr8", "100 moments")
- Removed MBTI strings
- Removed stopwords like "I" , "a", "the"

Text Processing

Before:

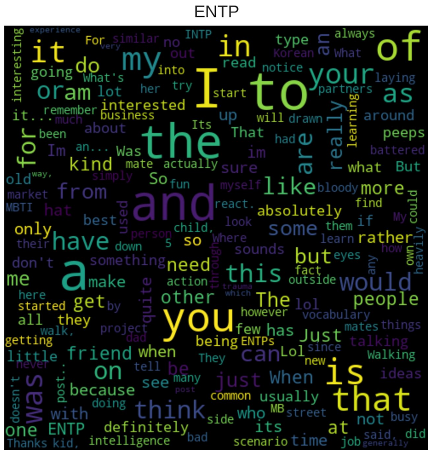
```
'https://www.youtube.com/watch?v=PLAaiKvHvZs|||51 :o|||I went through a break up some
months ago. We were together for 4 years and I had planned my life around that
relationship. I wasn't the one breaking the relationship as you might imagine and all
our...|||ENFJ Puns so many puns.|||....
```

After:

```
['went', 'break', 'months', 'ago', 'together', 'years', 'planned', 'life', 'around',
'relationship', 'one', 'breaking', 'relationship', 'might', 'imagine', 'puns',
'many', 'puns',...
```

Text Processing

Before:



After:

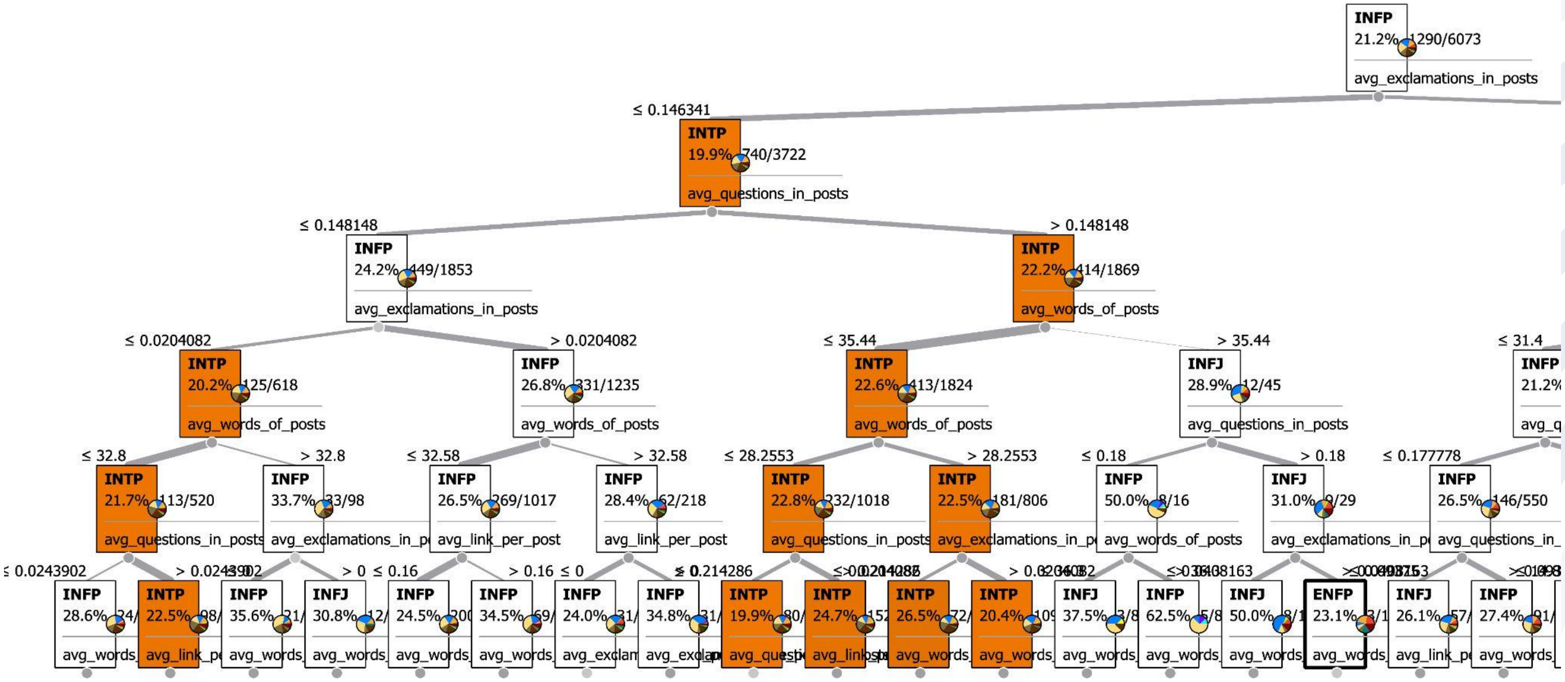
Project Process

Binary Tree

Data used for Binary Tree:

type	avg_link_per_post	avg_words_of_posts	avg_exclamations_in_posts	avg_questions_in_posts
INFJ	0,4800	16,1667	0,0833	0,0556
ENTP	0,2000	26,2128	0,0000	0,0851
INTP	0,1000	20,9286	0,0952	0,2143
INTJ	0,0400	22,6600	0,0600	0,1800
ENTJ	0,1200	22,5000	0,0217	0,1739
INTJ	0,0200	30,7800	0,0000	0,2000
INFJ	0,0400	28,4490	0,0612	0,2449
INTJ	0,0200	25,1200	0,0000	0,6800

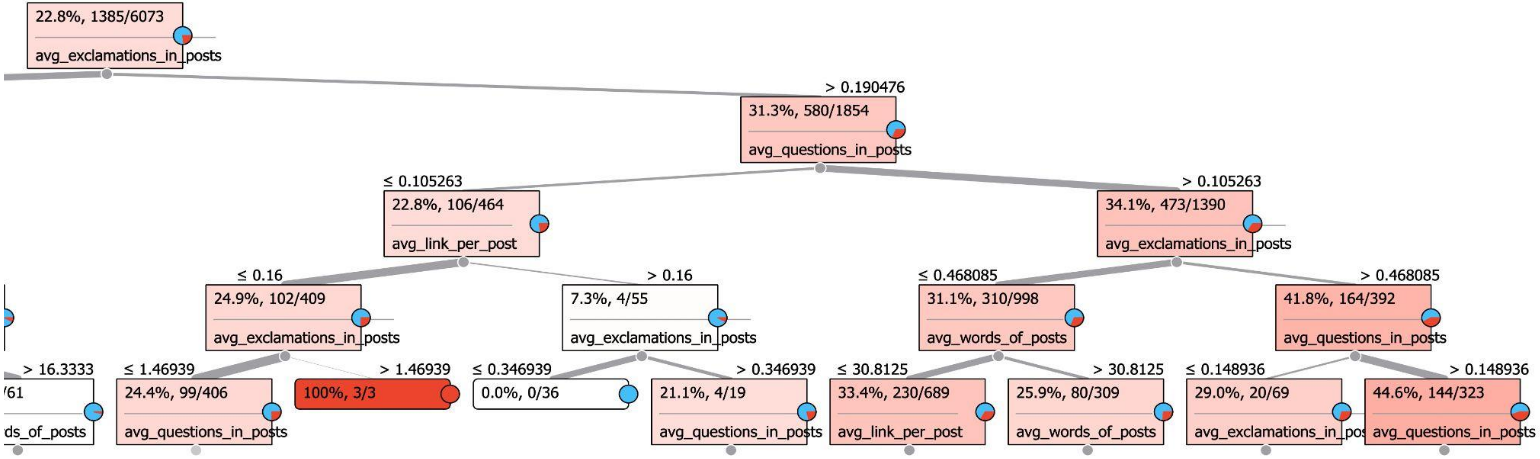
Binary Tree
exact prediction of the type



- Accuracy at 15,4%

Binary Tree

prediction Introvert or Extrovert



- Accuracy at 65%

One Hot Encoding

Steps:

1. Tokenize variables
2. Vectorize post to binary representation
3. Processing by Neural Network

Problem:

- Context of sentence is lost

TF-IDF : Definition

Weight:

TF * IDF = ~0.15

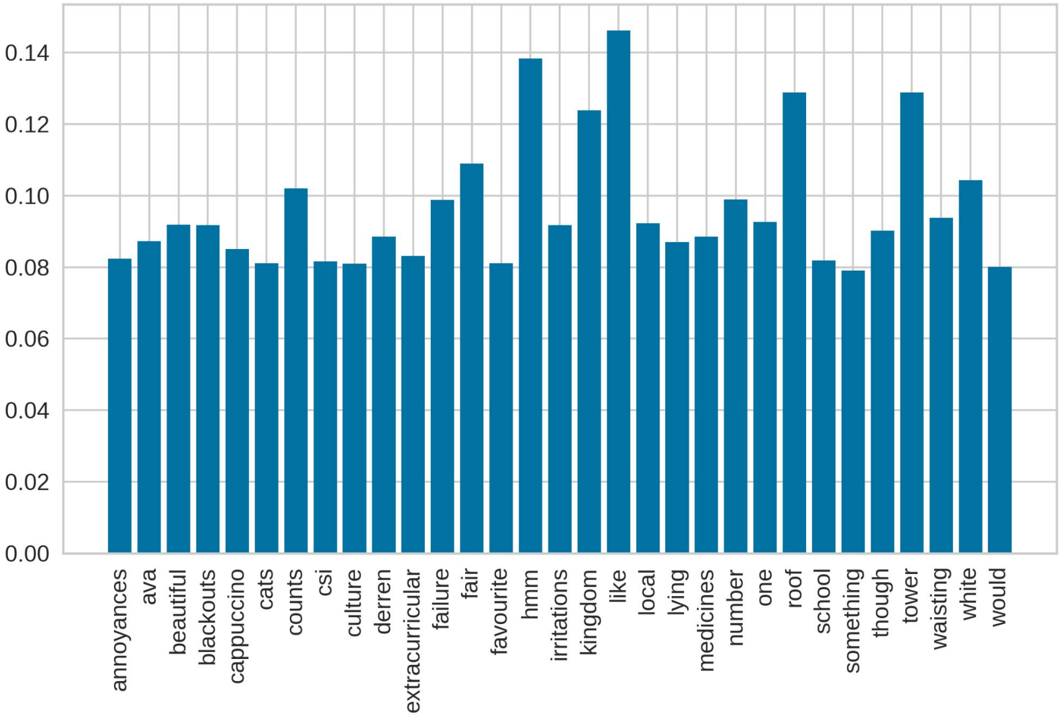
Term Frequency: (Frequency / Term Count)

TF(test) = 2 / 10 = 0.5

Inverse Document Frequency: (Document Count / Frequency)

IDF(test) = log(2/1) = ~0.3

TF-IDF : Transformation



N-Gram

(1,1)

['aba', 'abandonment', 'abbey', ... 'zwanglos', 'zylinder', 'zynthaxx']

(2,2)

['aba daba', 'ability able', 'ability change', ... 'zombie apocalypse', 'zone like', 'zooey deschanel']

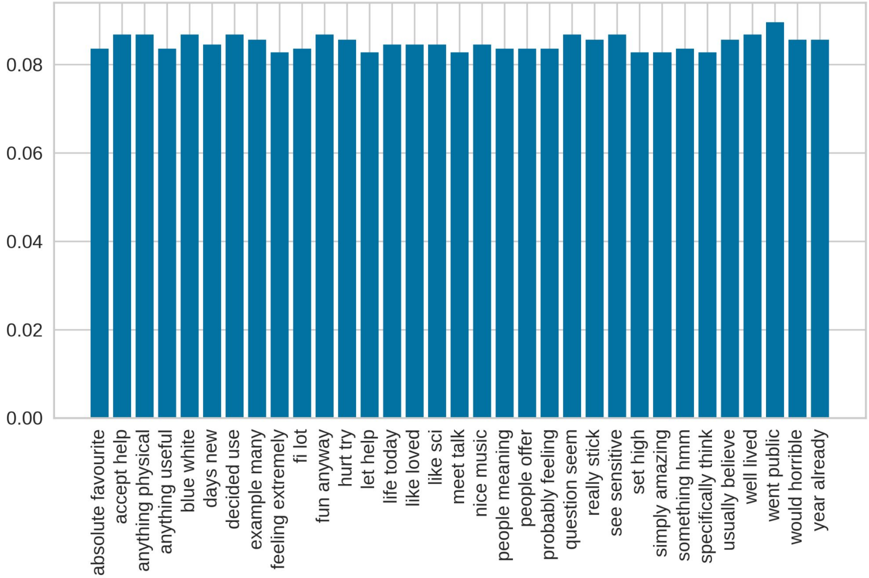
(3,3)

['aba daba aba', 'ability acquire apply', ... 'zero dark thirty', 'zero sum game']

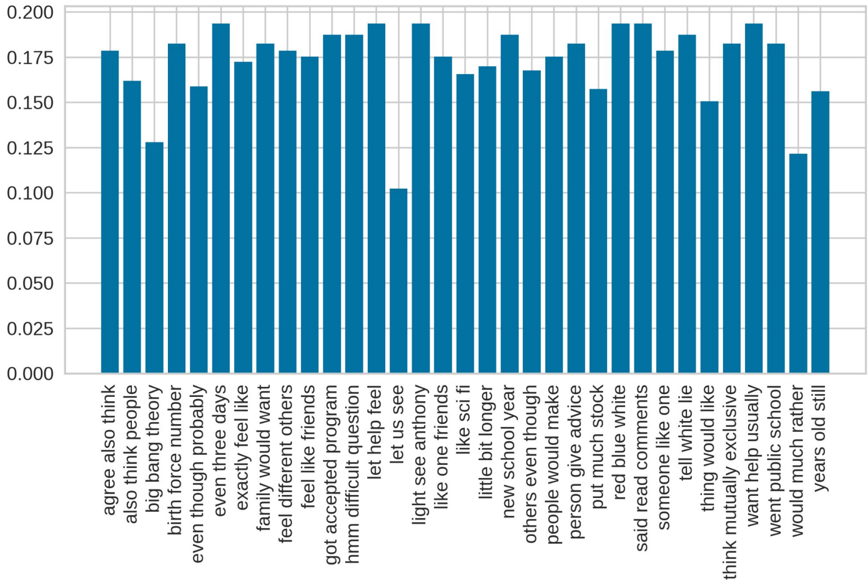
(1,2)

['ability', 'ability make', ... 'zodiac sign', 'zoe']

TF-IDF : N-Gram



2,2



3,3

MultiClass VS MultiLabel

I S T P

0 1 1 0 = 6

I S T P

[0 1 1 0]

Problem Transformation

Multi-Label $\Rightarrow$ Multi-Class

Naive-Bayes Classifier (NB)
Support Vector Machine (SVM)

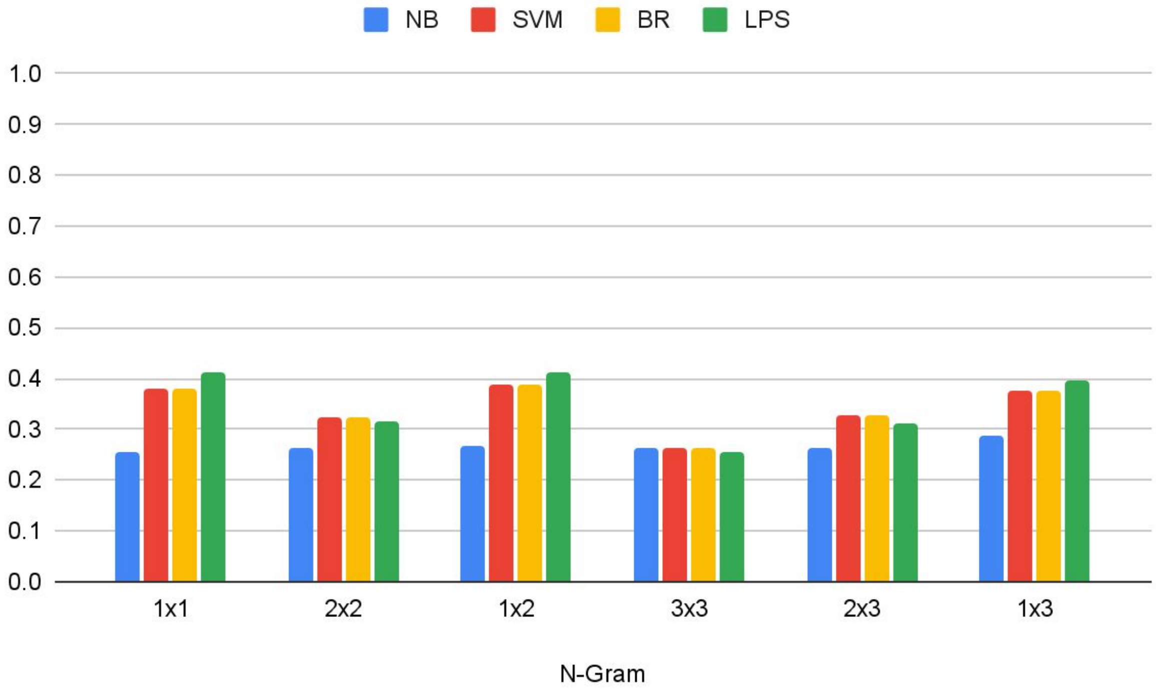
Binary Relevance (BR)
Label Power Set (LPS)

Label Power Set : Transformation

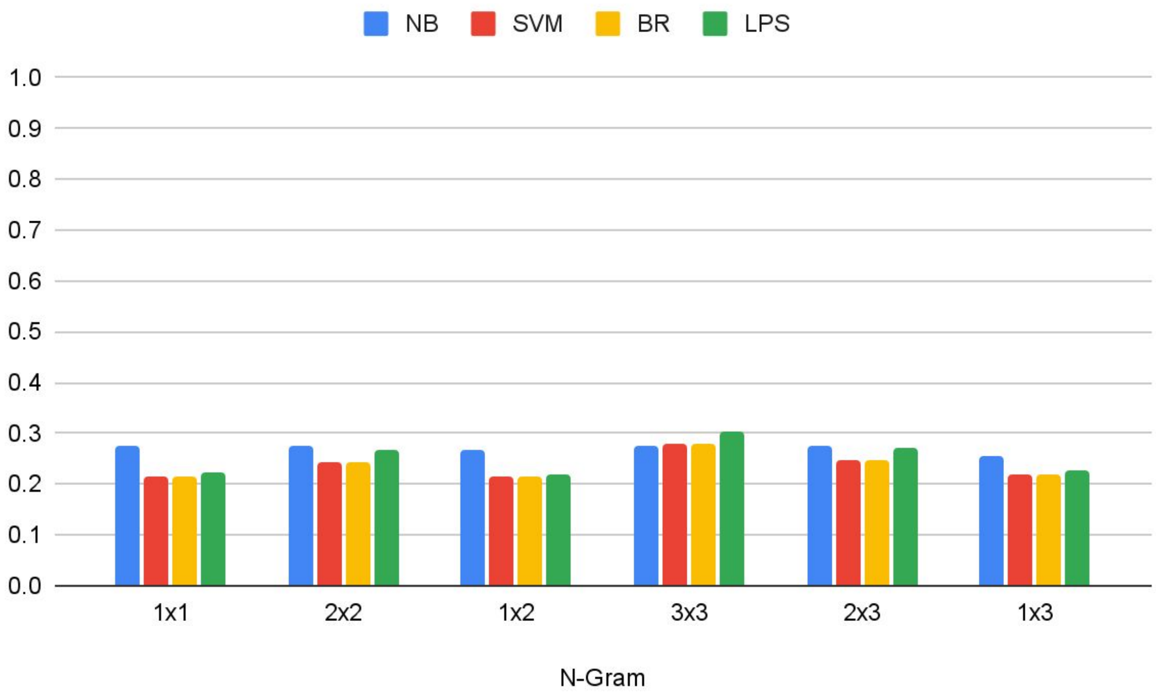
4 Labels $\rightarrow$ $2^4 = 16$ Classes

I S T P = [0 1 1 0] “I S T P” = 0 1 1 0 = 6

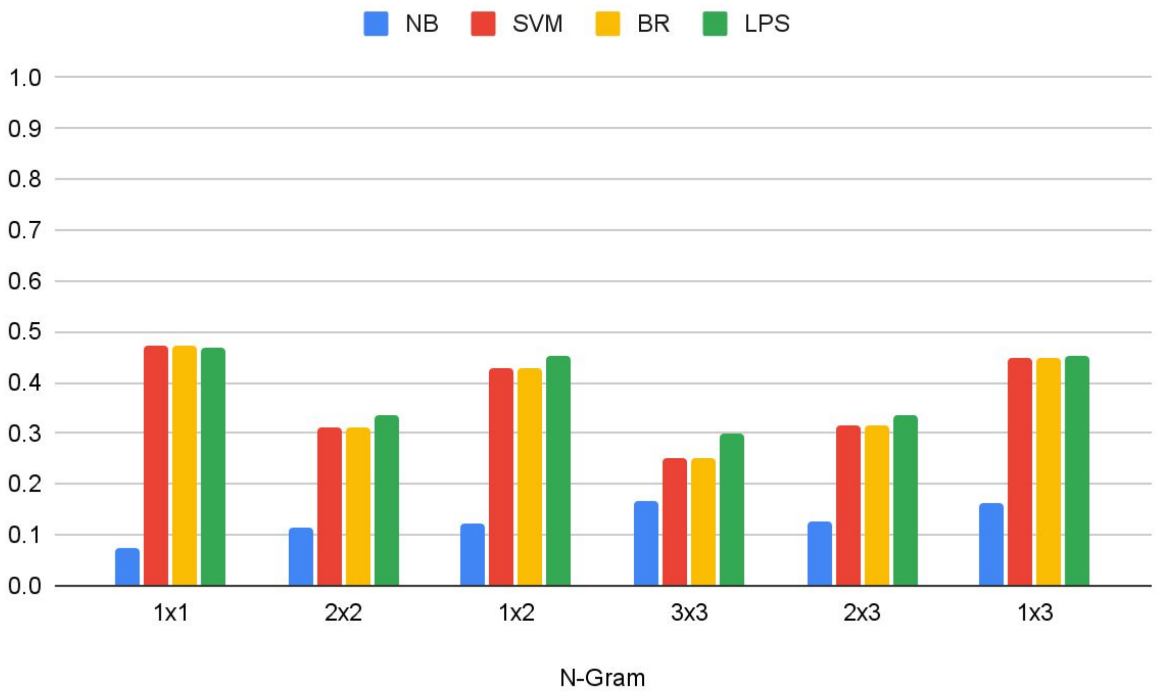
N-Gram : Accuracy Score



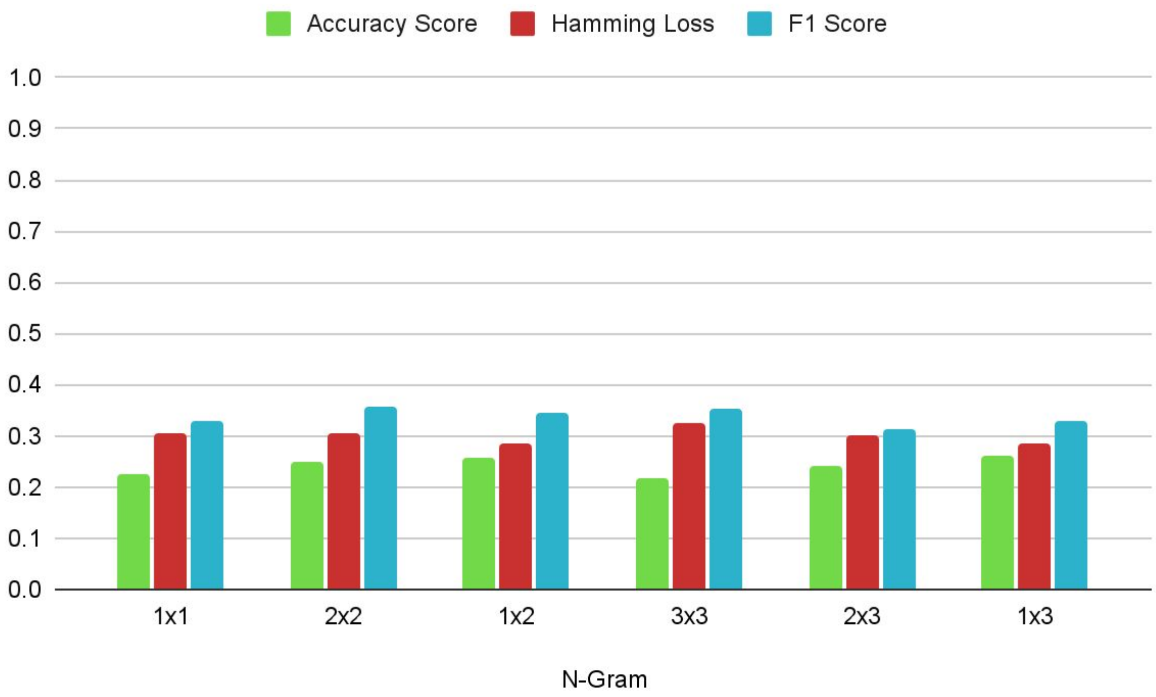
N-Gram : Hamming Loss



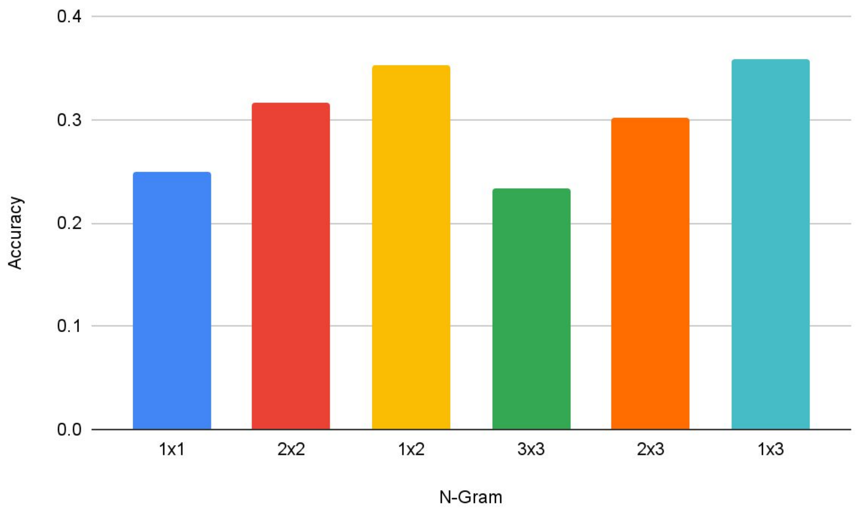
N-Gram : F1 Score

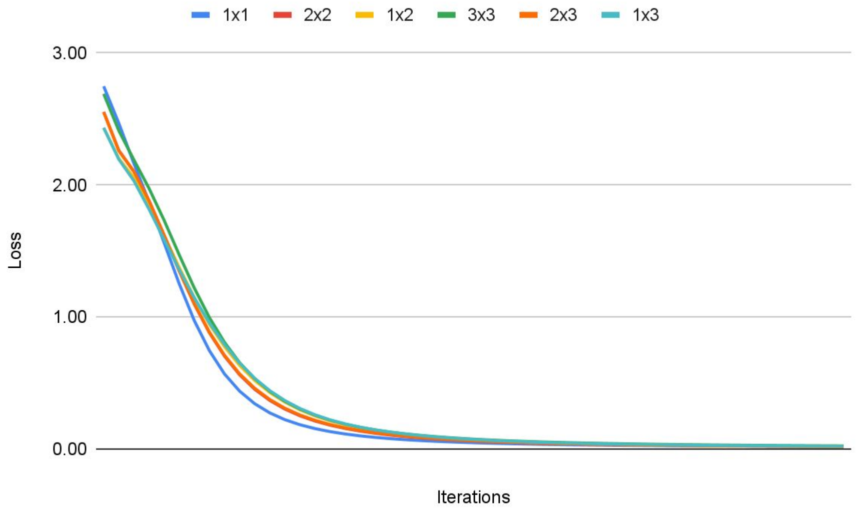


Adapted Algorithm : ML-kNN



Adapted Algorithm : ML-Perceptron





Conclusion

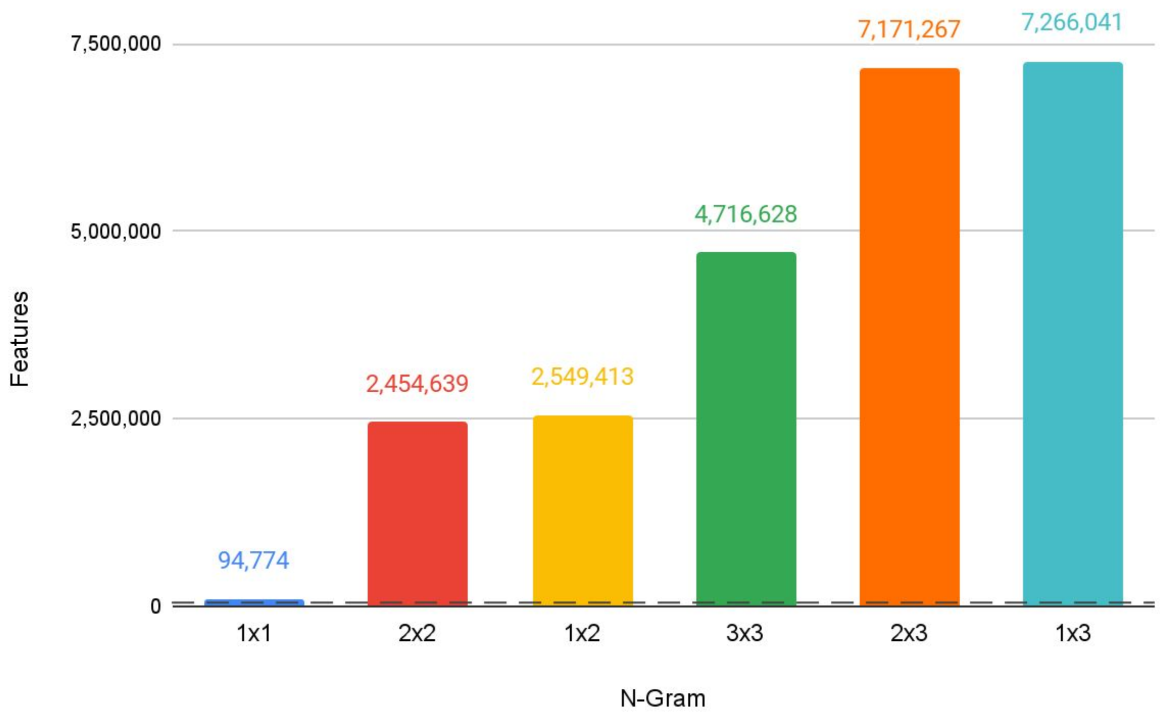
Results

- Random guessing : $50\%^4 = 6{,}25\%$
- Best guess 41,03%
    - LPS
    - (1,2)N-Gram
    - 50.000 features

→Better than guessing, but not reliable enough

Outlook

Feature Selection

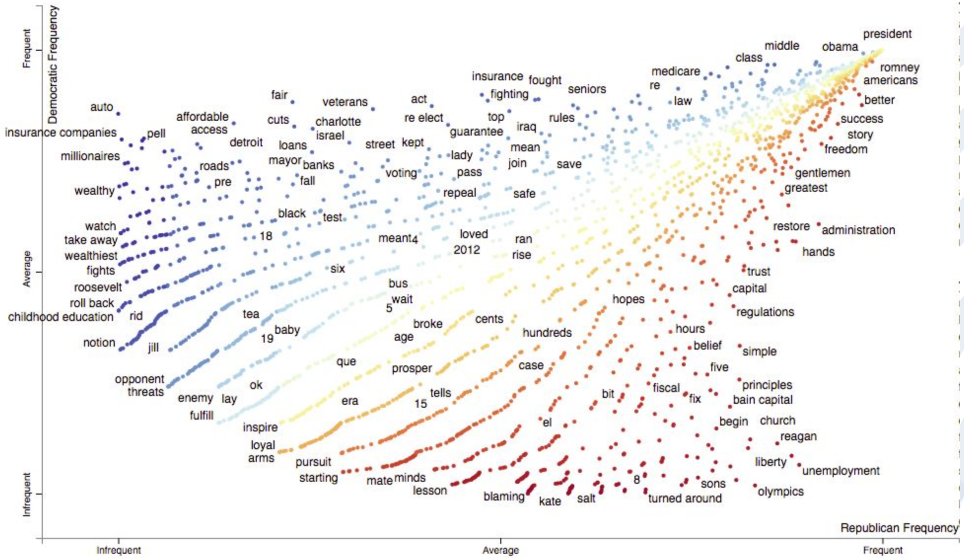


Vectorizer

Word2Vec:

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Bag-of-words:



Model Parameters

NB	LSVM	MLkNN	MLP
Smoothing	Penalty Norm	Amount of Neighbours	Size of Hidden Layers
Prior Class Probabilities	Loss Function	Smoothing	Activation Function
	Tolerance		Solver
	Regularization		Regularization
	Amount of Iterations		Batch Size
			Learning Rate

Sources

- https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/
- https://www.16personalities.com/articles/our-theory
- https://i.pinimg.com/736x/63/72/e4/6372e45c2749cf0f2863b7fd6f1759ba.jpg
- https://www.kaggle.com/aryansakhala/determining-personality-type-using-ml
- https://www.kaggle.com/prajwalkaushal/personality-prediction-mbti
- http://tfidf.com/
- https://www.researchgate.net/figure/3D-word-embeddings-visualization_fig2_338405739
- https://github.com/JasonKessler/scattertext
- https://medium.com/technovators/machine-learning-based-multi-label-text-classification-9a0e17f88bb4
- Efficient Estimation of Word Representations in Vector Space - Mikolov, Chen, Corrado, Dean
- Binary relevance for multi-label learning: an overview - Zhang, Li, Liu, Geng