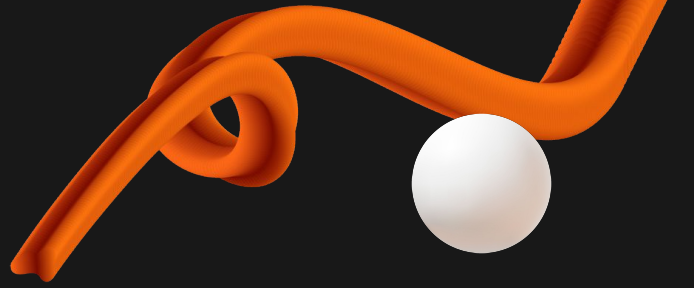
Influencer Sponsorship

Farah Liyana Binte Normal

Problem Statement



Employed by a marketing agency specializing on influencer marketing, we are tasked to perform market research on the social media platform.

- The client aims to reach a greater audience with the use of influencer marketing
- Creator should reflect the client's branding to boost overall brand recognition

Sample Clients

1. Brilliant

Brilliant is an online interactive classroom focus on STEM related subjects. The company aims to inspire and develop people in realising their STEM goals. They enable great teachers to illuminate the soul of math, science, and engineering through digestible content.

Key Words: Education, learning, teaching, STEM

Sample Clients

2. Nord VPN

NordVPN is a VPN service and the flagship product of the cybersecurity company Nord Security. We value people's freedom of choice beyond anything else, so we strive to offer access to free and safe internet for our users. Through hard work, dedication, and technological innovation we've created the fastest VPN in the world with state-of-the-art features.

Key Words: Lifestyle, cybersecurity, fullstack

Sample Clients

3. The Washi Shop

The company aims to build a community of artists, journalers, and crafters, we've designed and created an array original washi tapes and stationery for you to enjoy. Committed to quality of products and dedication to customers, The Washi Tape Shop creates the most vibrant gilded washi tapes with premium washi paper and eco-friendly printing technology around the clock to ensure end users has the best shopping experience.

Key Words: Stationary, lifestyle, washi

MODEL WORKFLOW

1

Data Scraping

YouTube Api documentation is used for scraping data

2

Problem Statement

Statement is fine-tuned to include clients' needs

3

EDA & Feature Engineering

Preliminary EDA and research for key indicators

MODEL WORKFLOW

4

Clustering

To find relations between existing text data this is included as a feature

5

Zero-shot Modelling

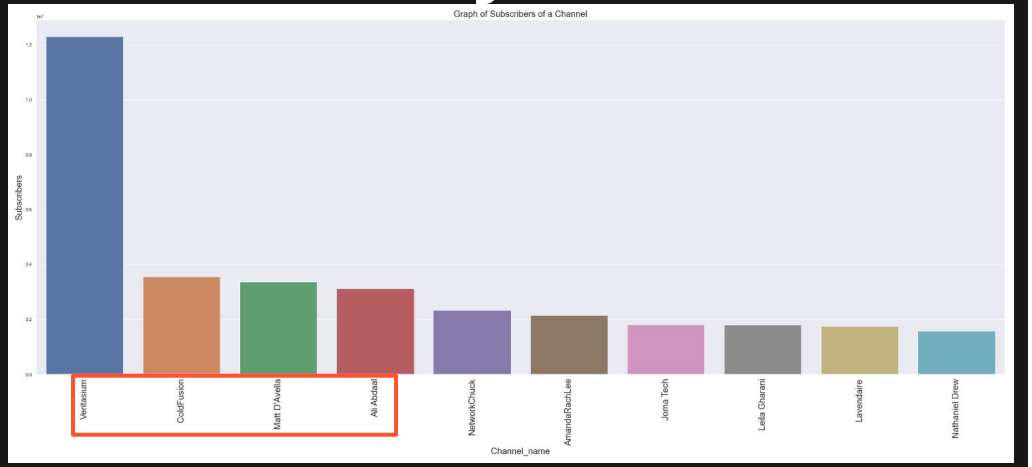
Labels are made based on the key words of the client

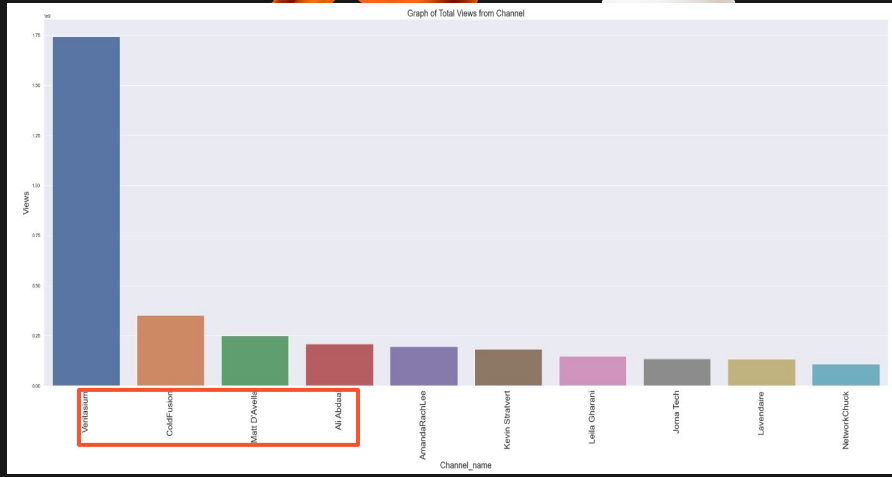
6

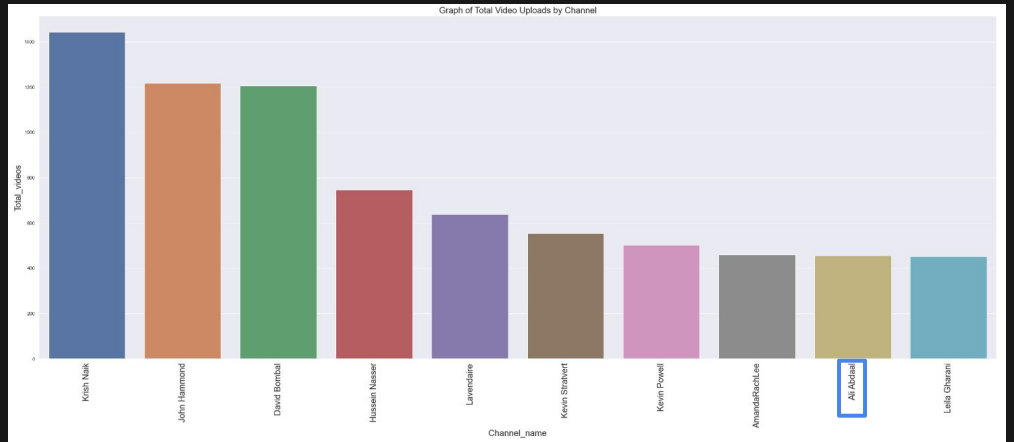
Results and Predictive Model

- Dataframe is merged on vid id and filtered for the creators.
- Pycaret is used for the prediction of target

Preliminary EDA



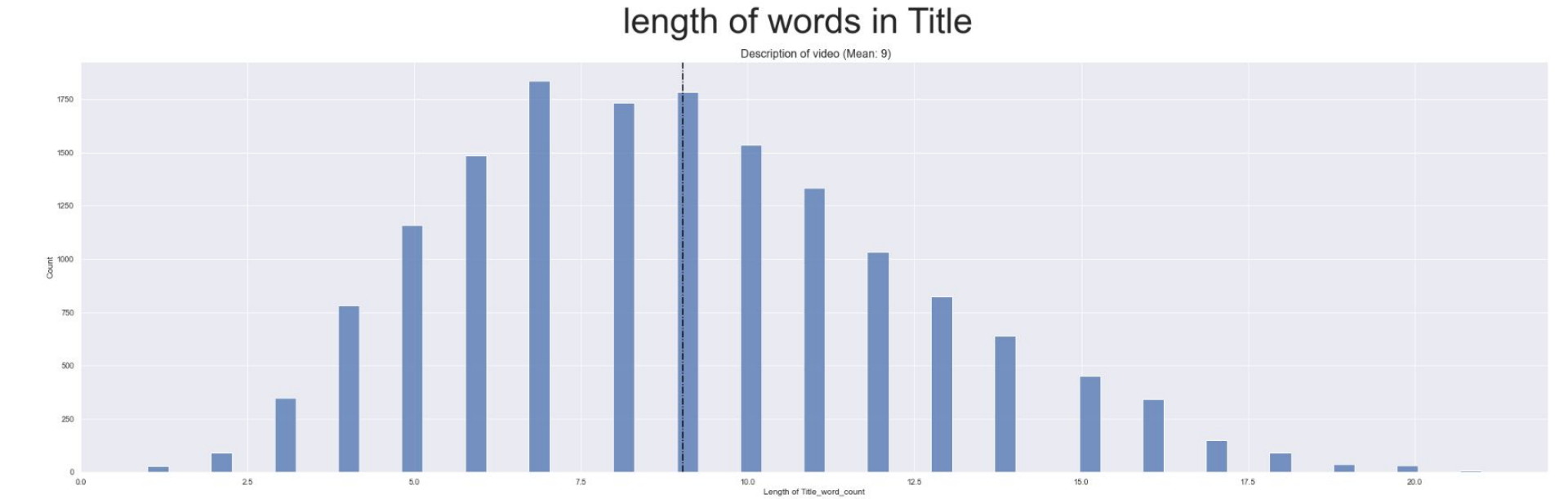




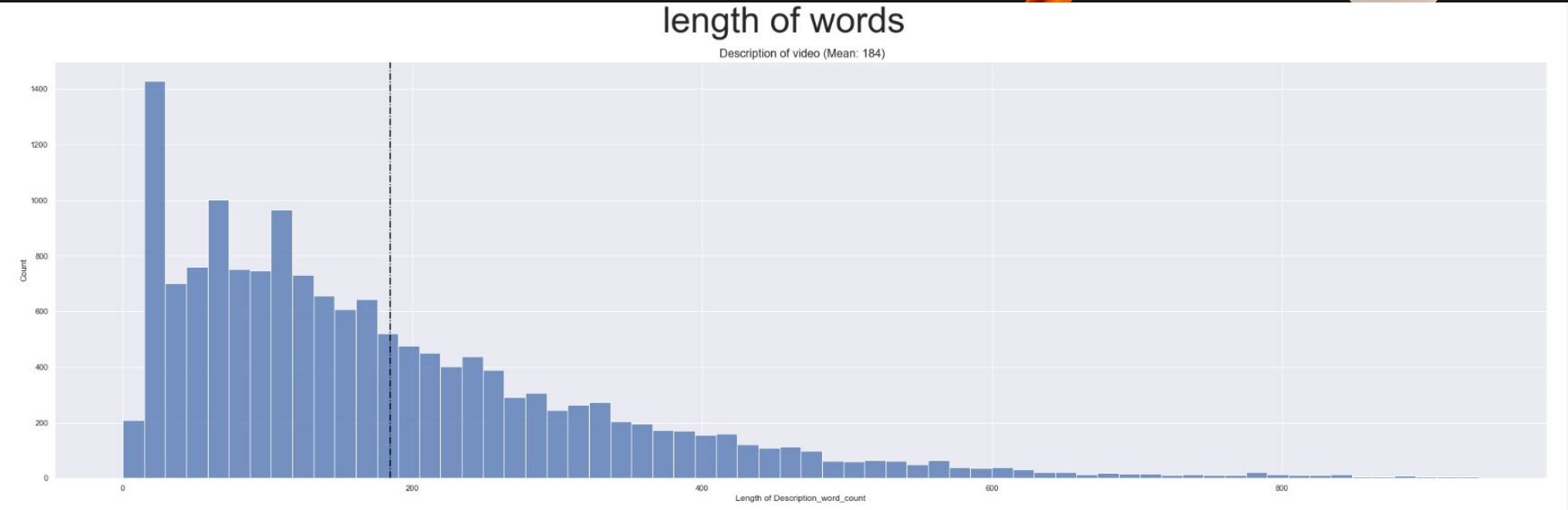
Summary:

- A stronger correlation can be found between the subscriber to the total channel views.
- weaker relationship exists between subscriber and views to the total video uploaded

Length of word in title



Length of words in text



Feature Engineering

Growth

An indicator of a channel growth has a 14% average ratio of view to subscribers.

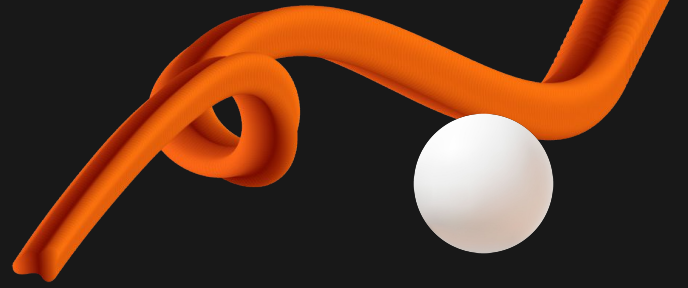
Engagement

The best engagement indicator for a successful video is 0.5% between the comments to view on an uploaded video.

Popularity

A benchmark of a good video is a video with 4 likes to every 100 view on an uploaded video.

Feature engineering



	Title	Channel_name	View_count	Like_count	Comment_count	View_per_day	Like_per_day	Comment_per_day	Engagement_ratio	Popular_target	Vid_id	Growth
0	Study with me videos are bad for actual learning	Tina Huang	16757.0	1113.0	143	9509.002236	631.587963	81.147420	0.853375	1	sS09Q-rKjos	36.992072
1	In Defense of Hustle Culture	Tina Huang	46006.0	2457.0	321	3555.394838	189.879692	24.807237	0.697735	1	A3x0gYW42Ts	36.992072
2	Why the best students are NOT the most successful	Tina Huang	77435.0	4551.0	379	2985.210750	175.446428	14.610898	0.489443	1	egBzplLuNhA	36.992072
3	How to Make a Self Study Plan (that you don't give up halfway)	Tina Huang	44254.0	2683.0	196	1306.302100	79.197554	5.785583	0.442898	1	SQNzij5m_Yg	36.992072
4	I left Meta...and I'm lowkey freaking out □	Tina Huang	207709.0	11628.0	799	4726.001495	264.571807	18.179642	0.384673	1	8tpha6Fla_k	36.992072

Numerical Feature Engineering

- View_count, Like_count and comment_count is averaged from uploaded date

Categorical Feature Engineering

- Use topic modelling on title
- Zero-shot labelling on the video title

Clustering Methods

Top2Vec

Top2Vec uses word embeddings in the algorithm. It is a continuous topic modeling method. The vectorization of text data makes it possible to locate semantically similar words within spatial proximity. Using pretrained embedding models such as the Universal Sentence Encoder.

StripNet

Stripnet runs on BERTopic model (continuous topic modelling) which is build on a similar to Top2Vec. The main difference between BERT to Top2Vec is the application of class-based term frequency inverse document frequency (c-TF-IDF) algorithm.

SpaCy & Gensim

Gensim and SpaCy is an library built on Latent Dirichlet Allocation(LDA) model. This is the most popular topic modeling technique and is a generative probabilistic model for discrete datasets such as text.

01. Top2Vec

```
topic_sizes_tit, topic_nums_tit = model_title.get_topic_sizes()
print(topic_sizes_tit[:5])
```

```
[669 442 390 351 351]
```

```
print(topic_nums_tit)
```

```
[  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89
  90  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107
 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125
 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143
 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161
 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
 180 181 182 183 184]
```

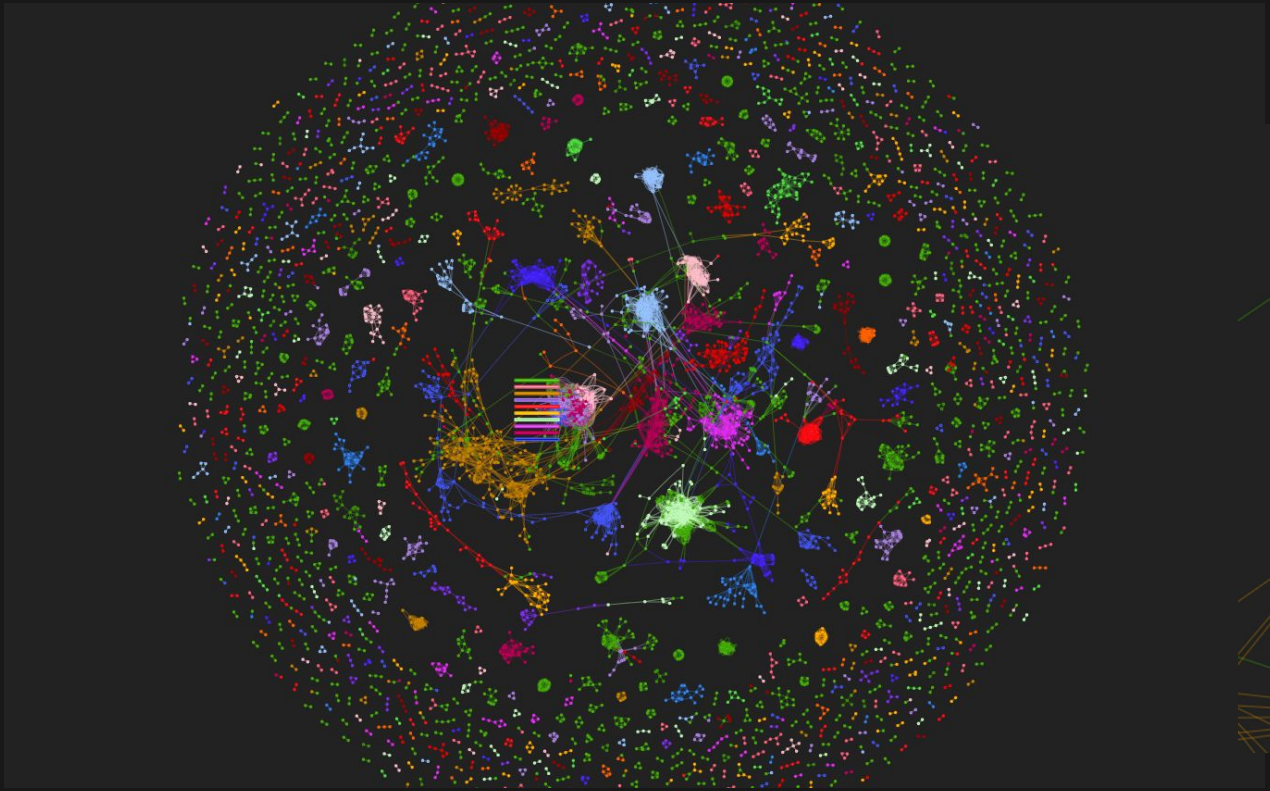
01. Top2Vec

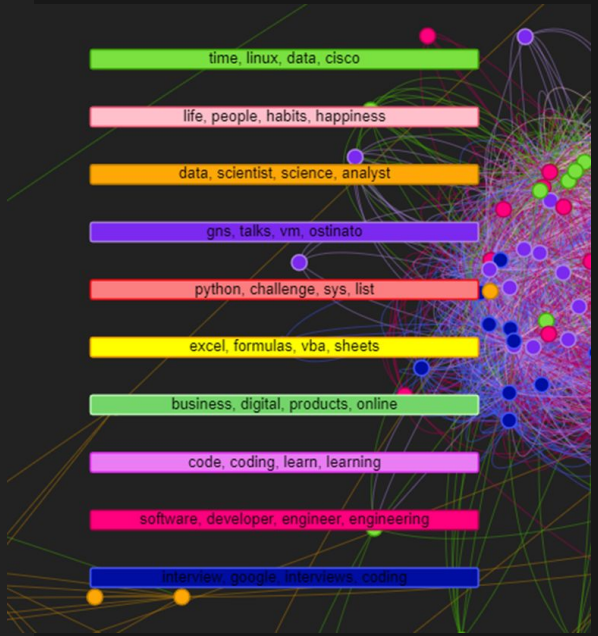
```
0
words : ['excel' 'charts' 'python' 'powerpoint' 'pivot' 'sql' 'javascript' 'chart'
 'database' 'microsoft' 'tutorial' 'analyst' 'grid' 'analytics'
 'programming' 'windows' 'tutorials' 'js' 'exam' 'api' 'html'
 'programmers' 'coding' 'data' 'tricks' 'resume' 'automation' 'vs' 'desk'
 'functions' 'programmer' 'productivity' 'beginners' 'tracer' 'github'
 'trick' 'hack' 'advanced' 'office' 'troubleshooting' 'django' 'function'
 'facebook' 'analysis' 'ultimate' 'fast' 'google' 'favorites' 'hacking'
 'css']
1
words : ['css' 'html' 'javascript' 'responsive' 'js' 'bgp' 'tutorials' 'web'
 'backend' 'python' 'coding' 'github' 'tutorial' 'react' 'django' 'ios'
 'website' 'beginner' 'iphone' 'sql' 'hack' 'programming' 'grid' 'docker'
 'beginners' 'programmers' 'text' 'hacking' 'facebook' 'setup'
 'powerpoint' 'ctf' 'creating' 'database' 'learning' 'ipad' 'custom'
 'cisco' 'design' 'excel' 'api' 'developer' 'troubleshooting' 'programmer'
 'ip' 'advanced' 'resume' 'charts' 'linux' 'windows']
2
words : ['python' 'learning' 'docker' 'django' 'ai' 'ccna' 'analytics' 'tutorial'
 'tutorials' 'ccnp' 'beginners' 'bgp' 'coding' 'database' 'programming'
 'github' 'hacking' 'api' 'ios' 'charts' 'programmers' 'troubleshooting'
 'wireshark' 'automation' 'learned' 'javascript' 'analyst' 'programmer'
 'data' 'advanced' 'backend' 'taught' 'excel' 'js' 'learn' 'networks'
 'scientist' 'beginner' 'machine' 'responsive' 'sdn' 'lessons' 'hack'
 'facebook' 'cisco' 'study' 'improve' 'batch' 'language' 'hpe']
3
words : ['gns' 'scientist' 'sparkles' 'bullet' 'tricks' 'crash' 'scratch'
 'science' 'fire' 'star' 'exam' 'trick' 'ground' 'linux' 'light' 'pivot'
 'ultimate' 'troubleshooting' 'mistakes' 'windows' 'apple' 'cloud'
 'powerpoint' 'tutorial' 'stationery' 'lab' 'macbook' 'engineers'
 'advanced' 'shorts' 'beginners' 'grid' 'hpe' 'life' 'hack' 'hacking'
 'tracer' 'tour' 'ctf' 'growth' 'programmer' 'challenge' 'api' 'quiz'
 'microsoft' 'iphone' 'engineer' 'programmers' 'javascript' 'subscribers']
4
words : ['python' 'django' 'javascript' 'html' 'sql' 'programming' 'js'
 'programmers' 'api' 'github' 'ios' 'programmer' 'backend' 'docker'
 'linux' 'coding' 'beginners' 'excel' 'css' 'hacking' 'beginner'
 'tutorial' 'iphone' 'database' 'windows' 'tutorials' 'bgp' 'wireshark'
 'hack' 'troubleshooting' 'computer' 'resume' 'ccna' 'ip' 'batch'
 'analyst' 'string' 'macbook' 'language' 'automation' 'hpe' 'amazon' 'ai'
 'pc' 'ccnp' 'microsoft' 'functions' 'mac' 'sdn' 'learning']
```

- data analyst
- software engineer
- data scientist
- hacking
- backend programming

02. StripNet

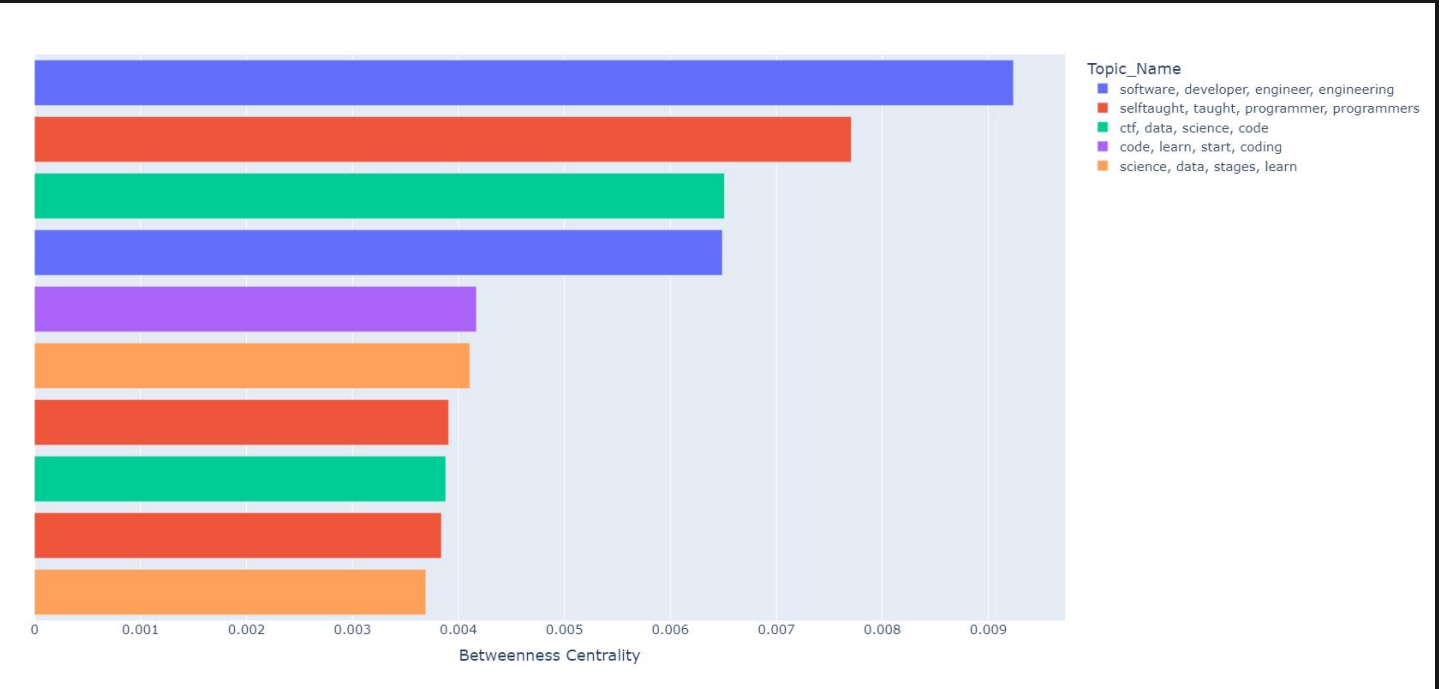
Best stripnet model on the text data tends to be below 0.8 threshold at 10 (min topics)





02. StripNet

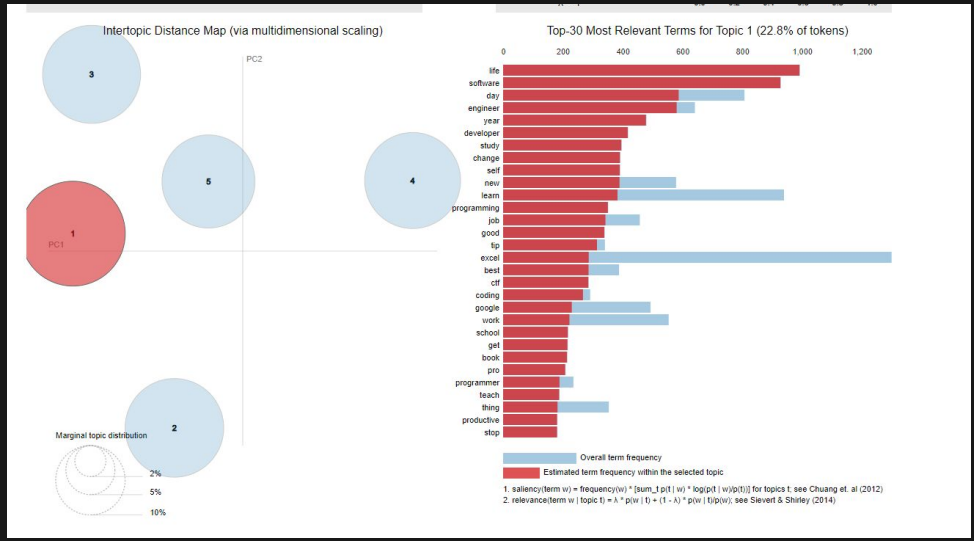
You can enter a subtitle here if you need it



03. SpaCy & Gensim

With the umass coherence score of -9.5, 5 topics is selected.

Topic No	Related topic
0	Fullstack
1	Learning
2	Coding
3	Business Tools (Tools)
4	STEM

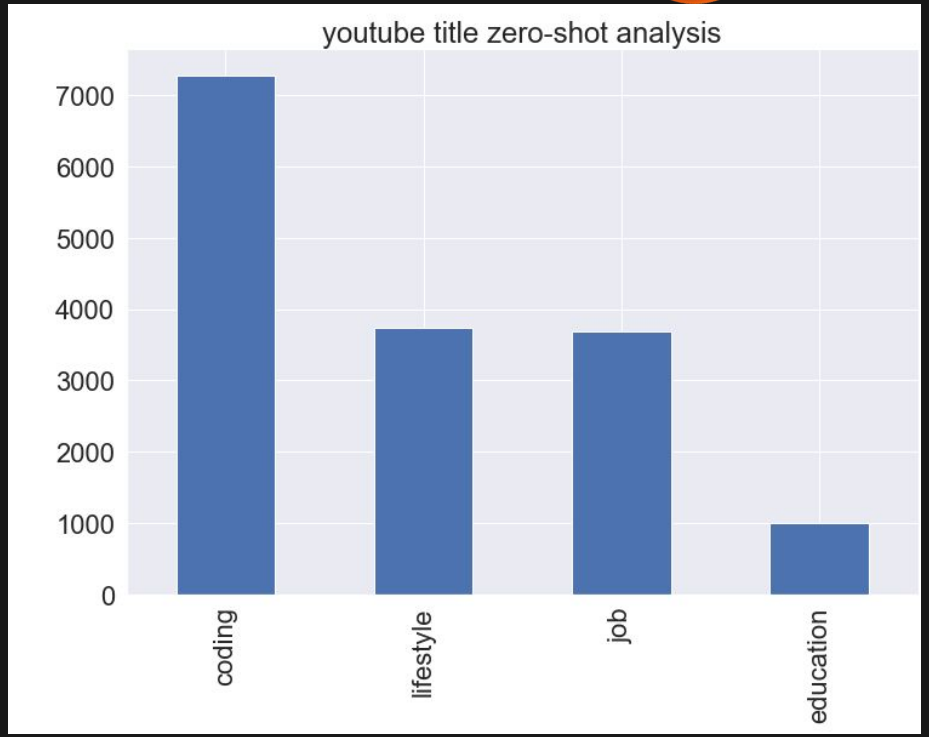



Zero-Shot
Modelling

Zero-Shot Modelling

Model Permutation	Accuracy
Bart-large-mnli (ML=F)	0.34
Cross-encoder/nli-distil roberta-base (ML=F)	0.55
Bart-large-mnli (ML=T)	0.26
Cross-encoder/nli-distil roberta-base (ML=F)	0.33

Accuracy based on 100 rows of data



Recommendation for Client

Client: Brilliant

Title	Channel_name	Vid_id	View_count	Like_count	Comment_count	View_per_day	Like_per_day	Comment_per_day	Engagement_ratio	Popular_ratio	Popular_target	Growth	zs_feature	topic
How to Study for the ACT and SAT (From a UChicago Grad) || ACE Your Tests and GET INTO COLLEGE!	The Almost Astrophysicist	lxFAXTblso4	270.0	32.0	12	0.377609	0.044754	0.016783	4.444444	11.851852	1	56.6625[illegible]6	education	learning
Fastest way to get hired as a self taught web developer	Dorian Develops	zTrQfrY3rrI	9079.0	1021.0	23	302.047404	33.967441	0.765182	0.253332	11.245732	1	22.0612[illegible]4	education	learning
2021 is gonna be LEET	John Hammond	beUAVHrepik	13527.0	1387.0	253	23.810496	2.441425	0.445336	1.870333	10.253567	1	4.6179[illegible]5	education	learning
What I learned from Posting on YouTube for Two years	Hussein Nasser	JHYd1q0-bhs	446.0	45.0	28	0.454107	0.045818	0.028509	6.278027	10.089686	1	7.2637[illegible]4	education	learning
holiday giveaway 🎁 school supplies & stationery!	studyquill	rko6p9K4iB0	73301.0	6773.0	1317	55.310259	5.110659	0.993760	1.796701	9.239983	1	20.7732[illegible]1	education	learning
Not Going to College: What I Wish I Knew	Nathaniel Drew	_nCImP9LMM	128429.0	11783.0	1266	683.420537	62.701915	6.736877	0.985759	9.174719	1	34.1639[illegible]0	education	learning
Teaching a KPOP IDOL BAND How to Draw! (The Rose)	AmandaRachLee	ocVDoAY-NiA	587655.0	53766.0	1635	640.259860	58.578948	1.781360	0.278224	9.149246	1	19.9897[illegible]0	education	learning
The Moment of Truth: Testing My Ability (Learning Portuguese part 2)	Nathaniel Drew	i8kAKcpYEEQ	375542.0	34292.0	3527	467.665157	42.704075	4.392198	0.939176	9.131336	1	34.1639[illegible]0	education	learning
How to Learn Anything | Learning Tips & Habits 🧠	Lavendaire	[illegible]zZGRBPxNmk	6610.0	599.0	31	302.645380	27.425807	1.419366	0.468986	9.062027	1	11.9958[illegible]8	education	learning
Advice & Tips For Self Taught Programmers (as a self taught programmer)	Dorian Develops	8[illegible]KOMAcoEQQ	24465.0	2126.0	148	58.937974	5.121689	0.356543	0.604946	8.689965	1	22.0612[illegible]4	education	learning

Creators to recommend to Brilliant are **The Almost Astrophysicist, Dorian Develops, John Hammond, Hussein Nasser** and **studyquill**.

Recommendation for Client

Client: NordVPN

Title	Channel_name	Vid_id	View_count	Like_count	Comment_count	View_per_day	Like_per_day	Comment_per_day	Engagement_ratio	Popular_ratio	Popular_target	Growt	zs_feature	topic
Thanksgiving giveaway	David Bombal	lIzIXmeFneU	13549.0	1943.0	1011	22.433528	3.217090	1.673946	7.461805	14.340542	1	4.17182	lifestyle	fullstack
60,000 Subscribers	ColdFusion	UeYMI5Rgrk	7257.0	954.0	215	2.288244	0.300811	0.067793	2.962657	13.145928	1	23.12710	lifestyle	fullstack
Massive 500K Giveaway	David Bombal	C9jxwi7HBnk	33902.0	4408.0	3643	68.770920	8.941721	7.389902	10.745679	13.002183	1	4.17182	lifestyle	fullstack
50K Subscribers - Celebration + Giveaway	Chandoo	4K36OSVIzF0	2860.0	349.0	327	3.622837	0.442087	0.414219	11.433566	12.202797	1	19.86744	lifestyle	fullstack
i've been to five BTS concerts & they all made me cry 🥲 #MyBTStory #ad	AmandaRachLee	JIgjATRzqRs	82339.0	9755.0	84	1887.199210	223.583336	1.925269	0.102017	11.847363	1	19.98976	lifestyle	fullstack
nourish yourself like a plant 🌿	Lavendaire	VLPJPp2qfYM	11920.0	1410.0	15	183.834838	21.745564	0.231336	0.125839	11.828859	1	11.99586	lifestyle	fullstack
Thankful For You	John Hammond	QtWmkIIKJY	2889.0	327.0	42	2.983135	0.337655	0.043369	1.453790	11.318795	1	4.61797	lifestyle	fullstack
Hacking Giveaway // Be Quick!	David Bombal	LntZCkPa_Ic	44209.0	4984.0	788	133.936696	15.099651	2.387345	1.782442	11.273723	1	4.17182	lifestyle	fullstack
Goodbye 2020	David Bombal	Lbtg23PpP6E	17029.0	1912.0	1401	29.929656	3.360473	2.462355	8.227142	11.227905	1	4.17182	lifestyle	fullstack
Christmas Giveaway Day 1	David Bombal	cEaKqQDwY	14841.0	1651.0	1735	25.592352	2.847044	2.991896	11.690587	11.124587	1	4.17182	lifestyle	fullstack

Creators to recommend to NordVPN are **David Bombal, ColdFusion, Chandoo, AmandaRachLee** and **Lavendaire**.

Recommendation for Client

Client: The Washi Shop

Title	Channel_name	Vid_id	View_count	Like_count	Comment_count	View_per_day	Like_per_day	Comment_per_day	Engagement_ratio	Popular_ratio	Popular_target	Growth	zs_feature	topic
habit stacking: easiest way to build a new habit! 💪 #shorts	Lavendaire	ZUogRJ5JkBw	8426.0	1007.0	20	2966.268149	354.501783	7.040750	0.237361	11.951104	1	11.995868	lifestyle	tools
How Brands Use Psychology to Make You Poor (from a marketer)	Sam Lui	IS2bjijhp-w	2194.0	222.0	29	9.549132	0.966229	0.126219	1.321787	10.118505	1	83.383577	lifestyle	tools
let's go outside & be in nature 🥰🧘✨	Lavendaire	6HX1-NsLooM	11122.0	1070.0	15	141.068794	13.571625	0.190256	0.134863	9.620572	1	11.995868	lifestyle	tools
go beyond small talk ✨ more honest conversations	Lavendaire	sR8ftzBieWk	12359.0	1185.0	22	160.839415	15.421531	0.286307	0.178003	9.588154	1	11.995868	lifestyle	tools
The Hidden Force That Controls You	Leon Hendrix	NgCSfspuYZY	59089.0	5601.0	425	143.157443	13.569782	1.029666	0.719254	9.478922	1	174.664975	lifestyle	tools
5 Lessons from Atomic Habits in Under 60 Seconds #SHORTS	Ali Abdaal	Dyliijbw_aQ	484112.0	45111.0	196	2262.599926	210.835809	0.916047	0.040485	9.318298	1	14.683795	lifestyle	tools
Let Your Pain Motivate You	Dennis Ivy	yngo848avVg	130105.0	11880.0	934	174.271235	15.912857	1.251061	0.717882	9.131086	1	44.480156	lifestyle	tools
Simple advice for creative people	Nathaniel Drew	S9LYW731Iqw	148259.0	13448.0	733	828.532770	75.153000	4.096308	0.494403	9.070613	1	34.163980	lifestyle	tools
How To Stay Motivated When You Feel Like Giving Up	Aurelius Tjin	-vHu_5RchtQ	901.0	81.0	34	1.634775	0.146966	0.061690	3.773583	8.990011	1	18.550677	lifestyle	tools
ways to level up your hydration 💧 #shorts	Lavendaire	RE5MTrZeC2c	26791.0	2314.0	26	225.908112	19.512201	0.219238	0.097043	8.637229	1	11.995868	lifestyle	tools

Creators to recommend to NordVPN are **Lavendaire, Sam Lui, leon Hendrix, Ali Abdaal** and **Dennis Ivy**.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
xgboost	Extreme Gradient Boosting	0.1220	0.0394	0.1984	0.8315	0.1407	0.1603	2.7760
lightgbm	Light Gradient Boosting Machine	0.1247	0.0410	0.2023	0.8247	0.1434	0.1661	0.1780
rf	Random Forest Regressor	0.1230	0.0458	0.2140	0.8040	0.1516	0.1651	3.2070
knn	K Neighbors Regressor	0.0985	0.0482	0.2192	0.7940	0.1508	0.1555	0.2000
et	Extra Trees Regressor	0.1807	0.0808	0.2841	0.6545	0.1997	0.2419	2.8590
gbr	Gradient Boosting Regressor	0.2407	0.0923	0.3038	0.6049	0.2155	0.3140	1.9160
catboost	CatBoost Regressor	0.0871	0.0264	0.1358	0.5868	0.0970	0.1146	2.5980
dt	Decision Tree Regressor	0.0991	0.0991	0.3146	0.5758	0.2181	0.1350	0.1020
ada	AdaBoost Regressor	0.3972	0.1781	0.4220	0.2376	0.3019	0.4779	0.4760
omp	Orthogonal Matching Pursuit	0.4058	0.2005	0.4477	0.1422	0.3117	0.5518	0.0320
llar	Lasso Least Angle Regression	0.4679	0.2340	0.4837	-0.0011	0.3403	0.6268	0.0330
dummy	Dummy Regressor	0.4679	0.2340	0.4837	-0.0011	0.3403	0.6268	0.0290

Problem Statement:

Positive outreach and greater outreach is the goal. Popular_ratio is the target.

Production Model: Lightgbm with and R2 score of 0.824.

Regression Modelling on Pycaret

Conclusion

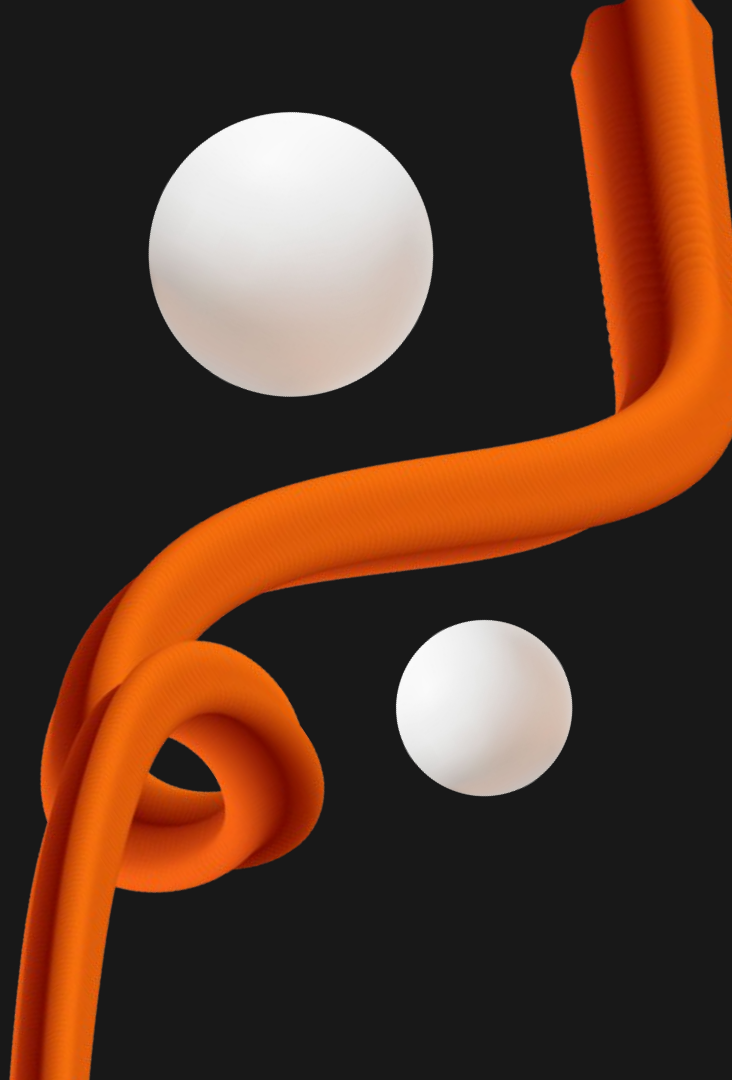
Companies who are interested in selecting content creators are able to match their company brand and creating more informed decisions for partnerships.

Recommendation

1. Increasing the dataset for social media by scraping.

2. Other models can considered for the clustering method

3. Increase labelled dataset.



Thanks!

Do you have any questions?

Email: farahliyananormal@yahoo.com.sg