

# Text Based Classification of Climbing Routes

General Assembly Capstone Project  
By: Ana Lincoln



# TABLE OF CONTENTS

**01** Background

---

**02** EDA

---

**03** Modeling

---

**04** Conclusions  
and  
Future Work

---



# 01

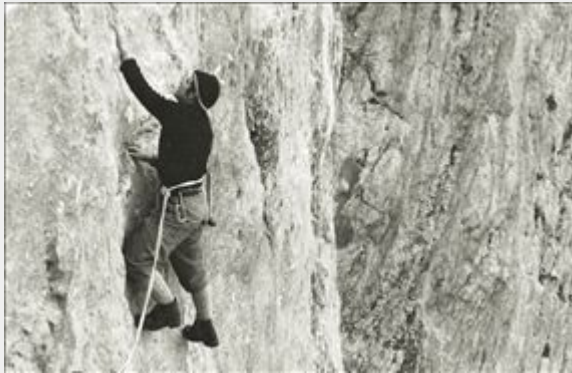
---

# Background



# What is Climbing?

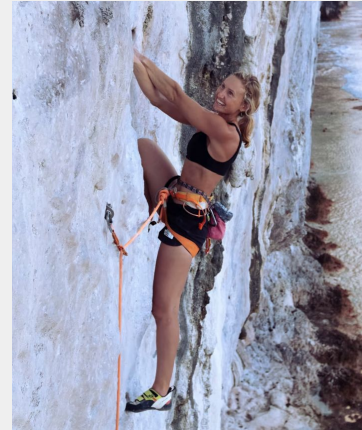
- Outdoor or indoor activity where you use specialized gear to ascend rock faces
- Gained massive popularity during the pandemic, especially outdoor climbing
- Routes most commonly defined by a description and a grade



[source](#)



[source](#)



[source](#)



# “Beta”

- The description of the location and difficulty of climbing routes
- Formerly kept solely in guidebooks (and by word of mouth)
- Climbing websites created new access to mass beta



# Problem Statement

**Given a climbing  
route's description,  
can we predict the  
grade?**



# Dataset



- Free, crowd-sourced beta repository with over 270k routes
- Formerly had an API to access data

All Locations > Utah > Southeast Utah > Castle Valley > Rectory


**Fine Jade** [🔗](#)

5.11a YDS ★★★★★ Avg: 3.9 from 587 votes

Type: Trad, 5 pitches, Grade III  
FA: Chip Chace & Pat Ellingwood, 1984  
Page Views: 79,728 total - 315/month  
Shared By: Ben Folsom on Oct 25, 2001 - Updates  
Admins: slim, Andrew Gram, Nathan Fisher, Perin Blanchard, GRK, DCrane

**You & This Route** 587 Opinions

Your To-Do List: [Add To-Do](#) - [View List](#)  
Your Star Rating: ★★★★★  
Your Difficulty Rating: -none- [Change](#)  
Your Ticks: [Add New Tick](#)  
-none-



⚠️ **Access Issue:** RAIN, WET ROCK and RAPTOR CLOSURES: The sandstone around Moab is fragile and is very easily damaged when it is wet. Also please ask and be aware of Raptor Closures in areas such as CAT WALL and RESERVOIR WALL in Indian Creek [Details](#) ▾

**Description** [🔗](#)

This is a classic desert line to the top of the rectory. Some great varied crack climbing. This is a must do desert climb. To reach, hike up the trail to the base of Castleton and then traverse North on the ridge for a couple hundred yards to reach the base of the rectory. The route climbs the crack system on the Southern prow of the rectory facing Castleton Tower.

P1- Climb a fist crack past a pod and make a difficult thin hands move right. Climb up to a ledge with anchors, pass them and continue thin hands to fingers to a ledge with bolts and belay. (5.10+)

P2- Continue up the crack to a finger crack through a bulge. Past the bulge climb a hand crack past a small roof to a belay stance. (5.11a)

P3- Make some lieback moves, then pass a short rotten section to a roof. Continue up the crack to a big ledge and belay or continue up the 5.11 bolted face to the top. If climbing the original route, traverse directly left from the big ledge below the top for 25-30 feet or so and top out a steep crack in a corner.

Descent- 3 double rope rappels down the route or 5 single rope rappels with a 70M

**Protection** [🔗](#)

2-3 sets of cams to #3 Camalot. A set of stoppers.



# Dataset



"An open source project that enables new and innovative uses of data about rock climbing routes." [Source](#)

Latest dataset has 127008 5th Class US Routes scraped from Mountain Project

	route_name	safety	fa	description	location	protection	mp_sector_id	mp_route_id	grade.YDS
0	Gravel Pit		Jason Milford/ Matt Schutz Spring 2020	[Goes up slab on bolts to steep corner on gear. Leaves the corner via steep roof traverse crack,...		[Chains on top, can lower off easy. Pro to 3" and a few draws for bolts.]	119029240	119029258	5.12b/c
1	Random Impulse		"Unknown" or	[Some fun moves broken up by a few scree filled ledges and a big bush. Crux comes half way up wh...	[25 feet to the right of Deep Springs Education.]	[A small assortment of cams and maybe a nut or two]	119100232	119101118	5.7
2	The Tick Wall		7, July 2020	[Bouldering. Approximately 14' tall and 20' or so wide that I have cleared and cleaned to date, ...	[Park at Sycamore Creek bridge and walk upstream along the old asphalt road for approximately 30...	[None. Bring your own pad.]	119181845	119181945	V-easy
3	Orange Crush		Wade Griffith, Sterling Killion, Scott Williams	[Pretty cool orange arete that sports some interesting climbing. Crimpy edges start you off cli...	[The route is located on the far southern shoulder of Yeti dome which is actually located on the...	[7 QD's]	105817198	105817201	5.11b/c





# Features

## Grade

- Difficulty rating of the route

## Description, Location, Protection

- Freeform text features that hold the beta for the route

## Climbing type

- Categorical feature describing the style of climbing

## FA - First Ascent

- Description of the first party to climb a route, sometimes including the year established



# YDS Grading System

## Class 5 sub-categories

5.1-5.4	Easy	A steep section that has large handholds and footholds. Suitable for beginners.
5.5-5.8	Intermediate	Small footholds and handholds. Low-angle to vertical terrain. Beginner to intermediate rock climbing skills required.
5.9-5.10	Hard	Technical and/or vertical, and may have overhangs. These hard climbs require specific climbing skills that most weekend climbers can attain.
5.11-5.12	Hard to Difficult	Technical and vertical, and may have overhangs with small holds. Dedicated climbers may reach this level with lots of practice.
5.13-5.15	Very Difficult	Strenuous climbing that's technical and vertical, and may have overhangs with small holds. These routes are for expert climbers who train regularly and have lots of natural ability.

[source](#)



# Data Cleaning

spaCy is a powerful NLP tool built for production that I used for:

- Lemmatization
- Stop word removal
- Exploring out-of-vocabulary (corpus-specific) terms

# spaCy



# Final Features

Feature	Type	Description
grade_reduced	int	The plain difficulty rating of the YDS grade (only 5th class routes were included)
lemmatized_text_combined	str	The combination of the three main text features, with various cleaned and lemmatized versions
type	bit	Actually 7 columns, each a one-hot representation of the climb type of the route
year_established	int	The year the route was established, extracted from the text of the FA



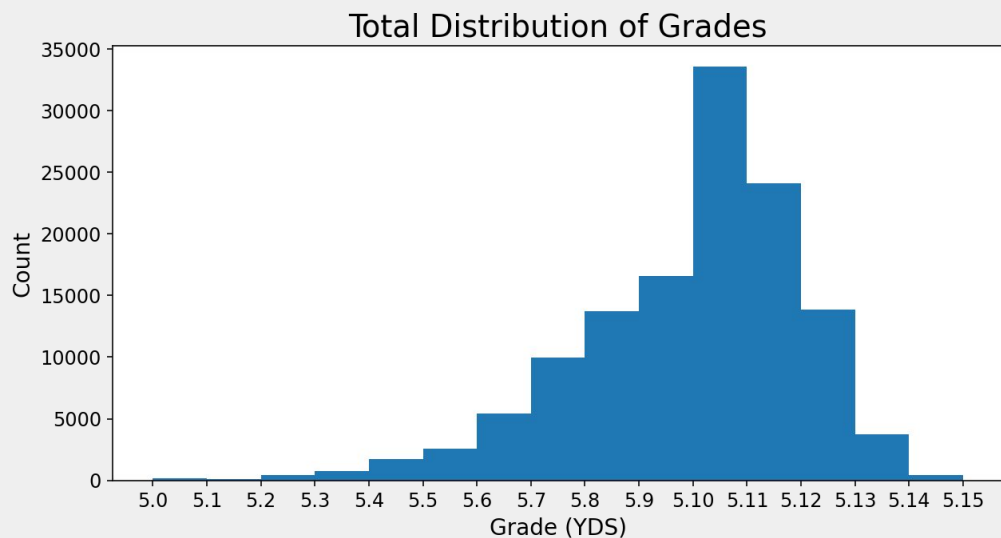
**02**

---

# Data Exploration



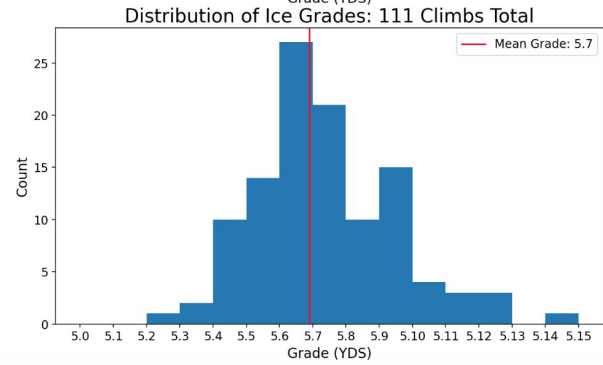
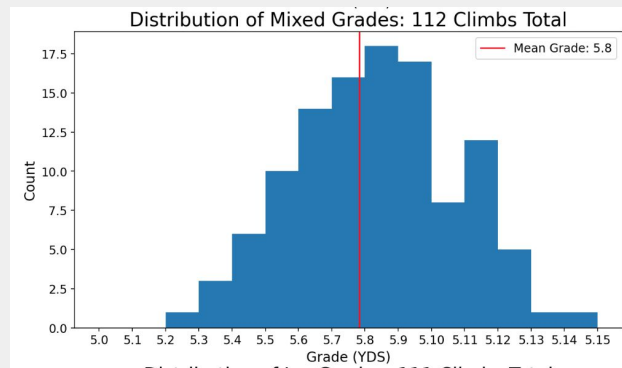
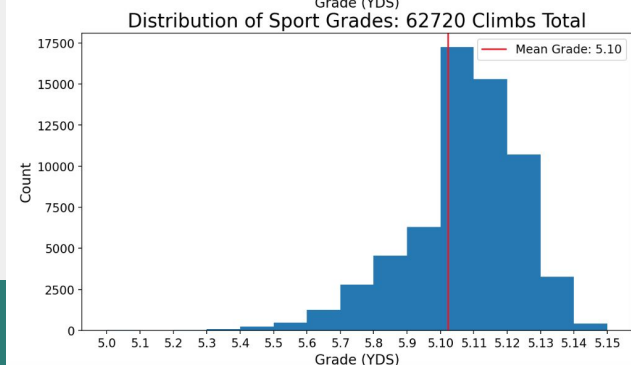
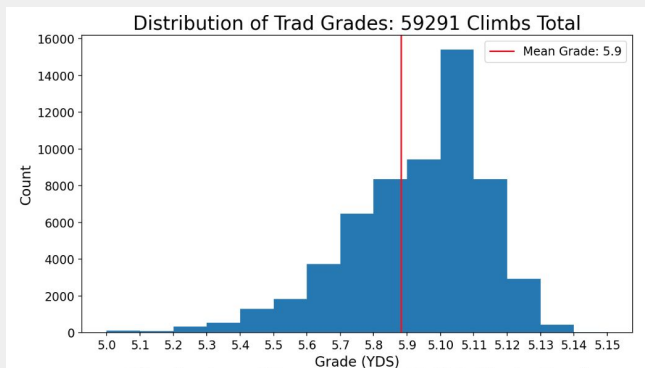
# Total Distribution of Grades



Grade	Num Rows	% Dataset
5.0	166	0.1
5.1	118	0.09
5.2	396	0.3
5.3	731	0.5
5.4	1,722	1.3
5.5	2,538	1.9
5.6	5,424	4.2
5.7	9,953	7.8
5.8	13,715	10.7
5.9	16,556	13.03
5.10	33,577	26.4
5.11	24,114	18.9
5.12	13,829	10.8
5.13	3,710	2.9
5.14	451	0.3
5.15	8	0.006

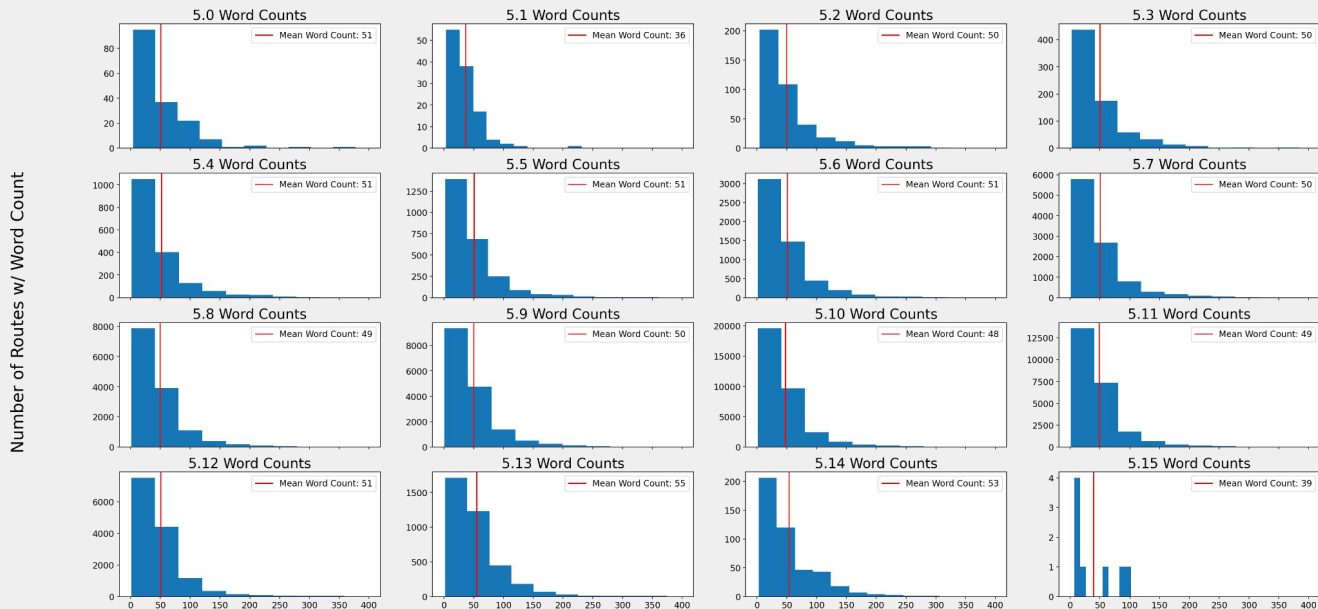


# Distribution of Grades by Type



# Word Counts

## Word Counts by Grade

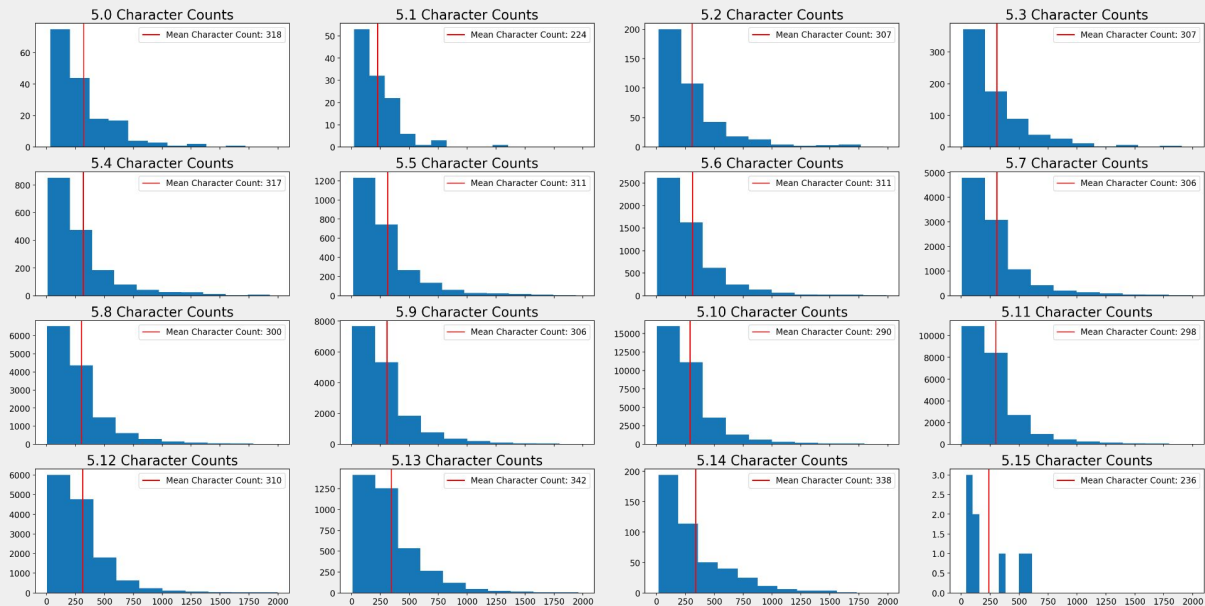




# Char Counts

## Character Counts by Grade

Number of Routes w/ Character Count



Combined Text Character Count



# Top TF-IDF Words

5.0-5.4:

- "easy" "scramble" "slab"

5.5-5.10:

- \* nothing notable \*

5.11-5.15:

- "roof" "crux" "hard" "problem" "project"

```
5.0 Top 10:
ridge      0.052026
rock       0.044617
route      0.044189
climb      0.042894
summit     0.042280
easy       0.040149
east       0.038290
face       0.036303
scramble   0.035364
right      0.034098
Name: 5.0, dtype: float64
```

```
5.1 Top 10:
crack      0.060904
route      0.049045
right      0.047954
climb      0.046514
tree       0.046408
easy       0.045032
rock       0.039998
gear       0.036956
face       0.036603
chimney    0.036599
Name: 5.1, dtype: float64
```

```
5.2 Top 10:
crack      0.055324
climb      0.054707
right      0.051932
easy       0.049305
route      0.044450
face       0.043247
slab       0.040020
rock       0.038203
left       0.036496
summit     0.034892
Name: 5.2, dtype: float64
```

```
5.3 Top 10:
right      0.058091
crack      0.056111
climb      0.055310
route      0.047852
tree       0.046485
face       0.044108
left       0.040378
corner     0.039240
easy       0.039041
chimney    0.038647
Name: 5.3, dtype: float64
```

```
5.4 Top 10:
climb      0.056693
right      0.056243
crack      0.055620
tree       0.045187
route      0.044605
face       0.043978
left       0.038899
anchor     0.036777
easy       0.035904
corner     0.034690
Name: 5.4, dtype: float64
```

```
5.5 Top 10:
crack      0.057030
climb      0.056542
right      0.055732
route      0.047101
tree       0.045444
face       0.043422
anchor     0.042441
bolt       0.041363
left       0.039970
leave      0.036275
Name: 5.5, dtype: float64
```

```
5.6 Top 10:
crack      0.061896
right      0.058371
climb      0.054988
bolt       0.049605
route      0.048366
anchor     0.047034
face       0.043718
left       0.038719
leave      0.036982
tree       0.035596
Name: 5.6, dtype: float64
```

```
5.7 Top 10:
crack      0.065596
bolt       0.060339
right      0.059752
climb      0.055797
anchor     0.051311
route      0.048347
face       0.046653
left       0.040190
leave      0.037147
start      0.033833
Name: 5.7, dtype: float64
```

```
5.8 Top 10:
bolt       0.069426
crack      0.065576
right      0.059963
climb      0.055319
anchor     0.052272
route      0.048815
face       0.046853
left       0.038241
leave      0.037341
start      0.035314
Name: 5.8, dtype: float64
```

```
5.9 Top 10:
bolt       0.077275
crack      0.061676
right      0.059777
climb      0.054544
anchor     0.052832
route      0.049667
face       0.046957
leave      0.038803
left       0.037270
start      0.036677
Name: 5.9, dtype: float64
```

```
5.10 Top 10:
bolt       0.092754
right      0.057071
anchor     0.053839
climb      0.051120
route      0.050994
crack      0.050457
face       0.046090
start      0.037827
leave      0.036900
left       0.034891
Name: 5.10, dtype: float64
```

```
5.11 Top 10:
bolt       0.097751
right      0.053904
route      0.050791
anchor     0.050442
climb      0.046071
face       0.040880
start      0.038502
crack      0.036929
leave      0.036317
roof       0.036130
Name: 5.11, dtype: float64
```

```
5.12 Top 10:
bolt       0.098610
route      0.052051
right      0.048762
anchor     0.047091
climb      0.040802
crux       0.038205
start      0.037047
roof       0.036890
leave      0.034094
face       0.030924
Name: 5.12, dtype: float64
```

```
5.13 Top 10:
bolt       0.088323
route      0.049356
right      0.042196
crux       0.041183
anchor     0.039123
climb      0.038086
start      0.037915
hard       0.036640
roof       0.032843
leave      0.032416
Name: 5.13, dtype: float64
```

```
5.14 Top 10:
bolt       0.074404
hard       0.041090
crux       0.039683
start      0.039014
climb      0.038575
route      0.038484
draw       0.034233
boulder    0.033717
problem    0.032898
right      0.031843
Name: 5.14, dtype: float64
```

```
5.15 Top 10:
project    0.196619
range      0.107681
ii         0.081594
ba         0.067813
jaws       0.065863
electric   0.062879
fence      0.061064
suspect    0.060008
somebody   0.057378
rasmussen  0.056525
Name: 5.15, dtype: float64
```



# 03

---

## Modeling



# Models

**BERT**

transfer learning

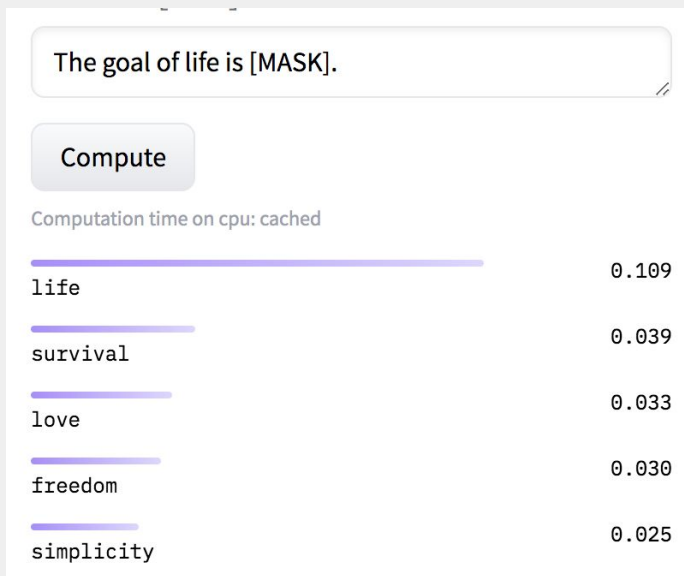
**fastText**

word2vec with  
multinomial logistic  
regression



# BERT

- Bidirectional Encoder Representations from Transformers
- Breakthrough NLP method of the early-mid 2010s
- Masked Language Modeling

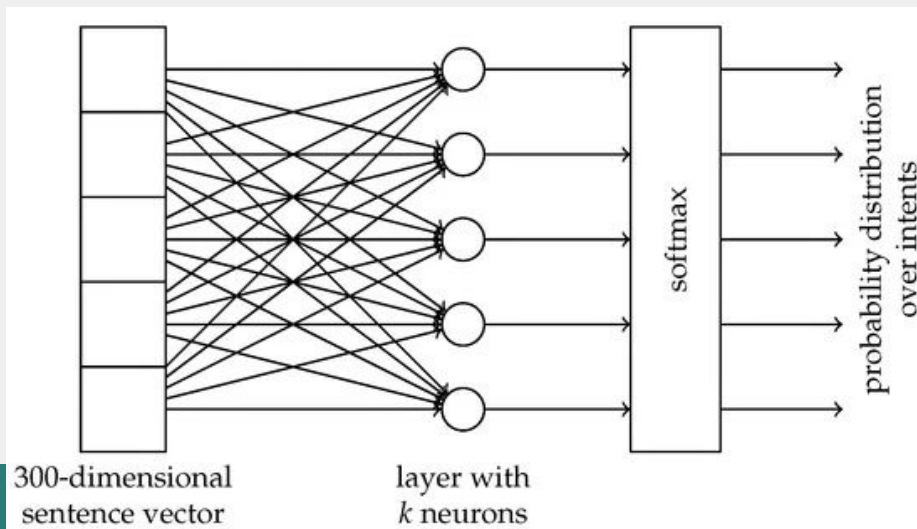


[source](#)



# fastText

- Optimized classifier developed by Facebook
- Single-layer neural network embedding + multinomial logistic regression with softmax output
- Lightning fast, similar performance to BERT



[source](#)



# Null Model

Predict only the majority class

- Null Accuracy: 26.4%
- Null Mean Absolute Error: 1.5



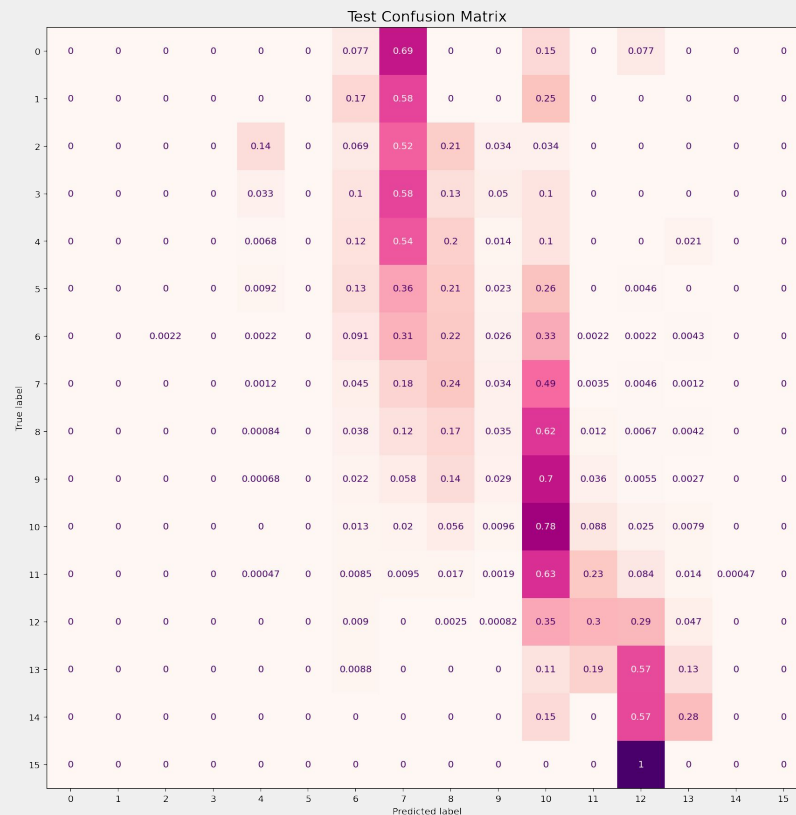
# Data Cleaning Iterations

- Removal of Corpus-Specific Stop Words
- Addition of Climb Type
- Spell Correction
- Reduction of Rows





# Post-Cleaning Confusion Matrices

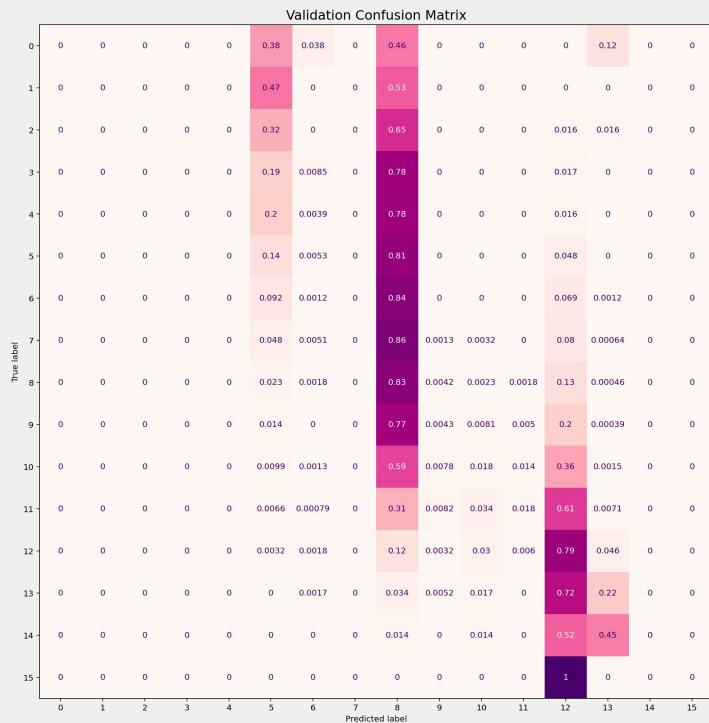


# Post-Cleaning Confusion Matrices

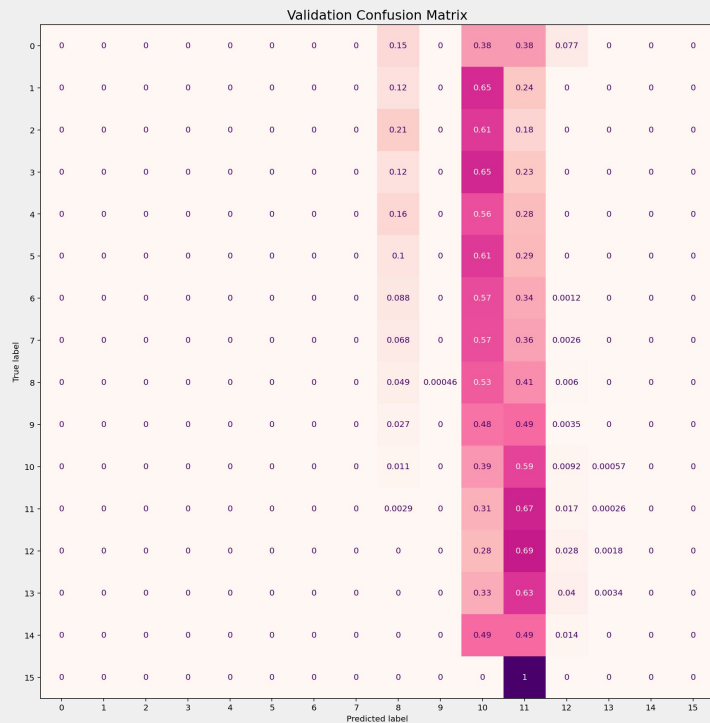


# Random Resampling

## Random Over Sampling



## Random Under Sampling

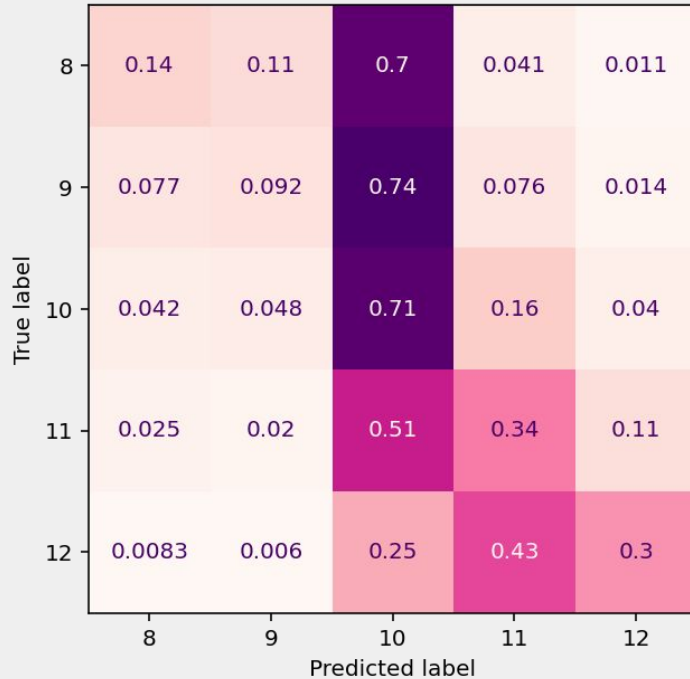


# Reduction of Classes

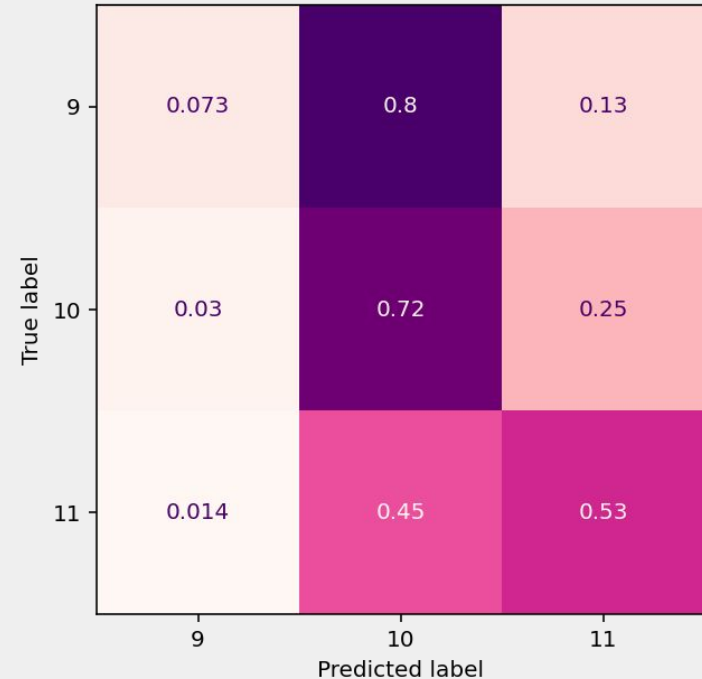
5.8-5.12

5.9-5.11

Validation Confusion Matrix



Validation Confusion Matrix



# Grouping of Classes

Group	Grade	% Dataset
0	5.0-5.8	27%
1	5.9-5.11	58%
2	5.12-5.15	14%

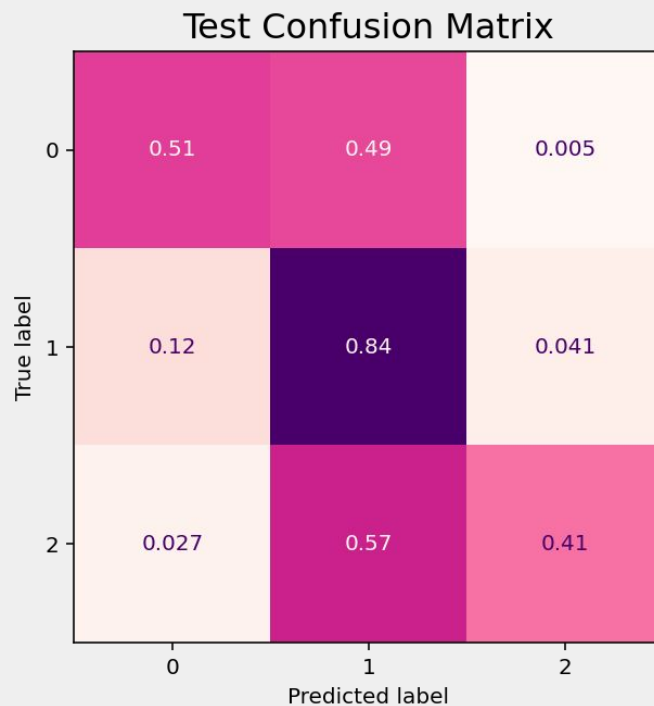
Null Accuracy: 58%

Null MAE: .41



# Grouping of Classes

Train Accuracy:	99.6%
Validation Accuracy:	68.3%
Test Accuracy:	68.8%
Test MAE:	0.32



# 04

---

## Conclusions and Future Work



# Conclusions

The Question:

**Given a climbing  
route's description,  
can we predict the  
grade?**





# Conclusions

The Answer:

**No**



# Future Work

- Document similarity using doc2vec, spacy
  - Deeper EDA
  - Potentially useful for resampling
- Further exploration of transfer learning techniques
- Introduction of ensemble methods



# THANKS!

Does anyone have any questions?



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

