

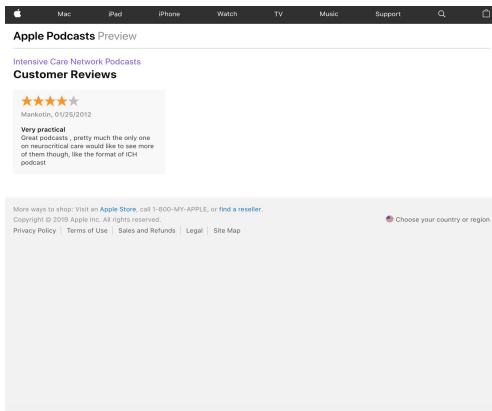
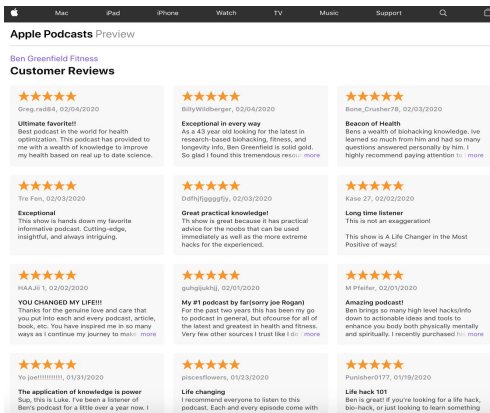
# Podcast Prophet

Predict podcast rating, provide suggestions based on the podcast reviews

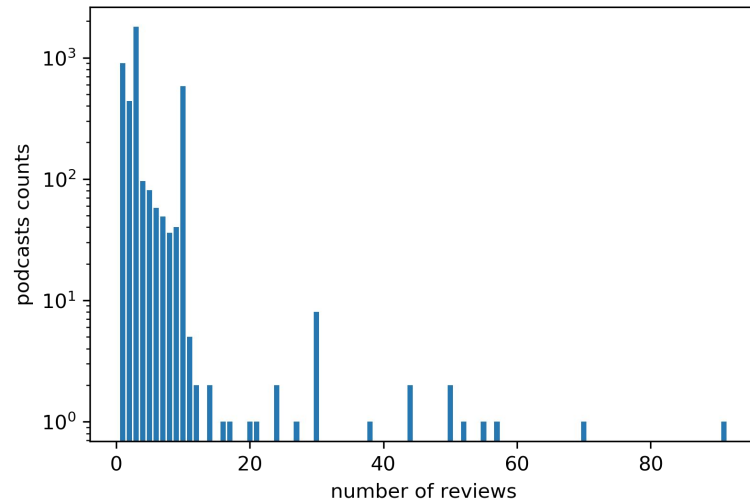
Ziying Feng



# Problem



Number of Reviews Distribution



- Among **10k+** fitness and health podcasts, **4115** podcasts have reviews
- Listeners: if the podcast has one review, is it a good podcast to listen to?
- Creators: how can I get more people to listen to my podcasts?
- Predict podcast ratings and provide suggestions based on reviews

# Web app

<http://datascienceanalysis.xyz/>



**PODCAST** *Prophet*

Please Enter Fitness Podcast Link:

**Submit**

Please input a fitness podcast link. More links at [itunes Preview](#).

Data



# Scraped 10k Fitness Podcasts from iTunes

Review  
Categorization

Description

Apple Podcasts Preview



## The Nutrition Diva's Quick and Dirty Tips for Eating Well and Feeling Fabulous

Quick and Dirty Tips

Nutrition

[Listen on Apple Podcasts](#)

JAN 28, 2020

### 465 - Can You Absorb Nutrients Through Your Skin?

With transdermal nutrition, you slap on a patch and don't need that vitamin reminder. But do these products really deliver or are they a waste of money? Read the transcript at <https://www.quickanddirtytips.com/health-fitness/trends-fads/can-you-absorb-nutrients-through-your-skin/>

▶ PLAY 9 min

300 episodes

Nutrition Diva Monica Reinagel serves up simple, painless ways to upgrade your eating habits. Before you know it, you'll be eating healthier and feeling more fabulous!

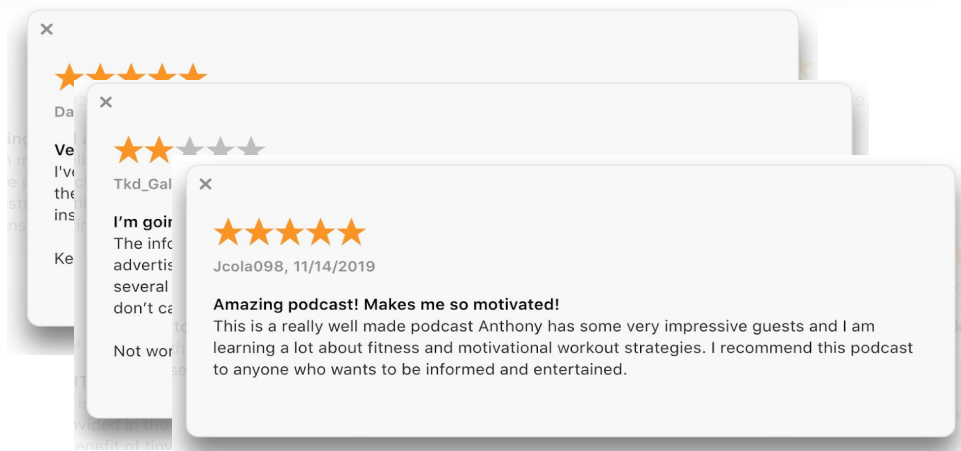
JAN 21, 2020

### 558 - Best and Worst Diets of 2020 with Jill Weisenberger

Registered dietitian and nutritionist Jill Weisenberger joins Nutrition Diva to share the results and an inside look at the process behind US News and World Report's annual Best Diet Rankings. TRANSCRIPT: <https://quickanddirtytips.com/health-fitness/healthy-eating/best-diets/>

▶ PLAY 23 min

Review



Sentiment Analysis

Description  
Vectorization

Prediction

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity spaCy

"Makes me so motivated! It has a fantastic and enormous variety of sounds and speeds. Very informative and gained some new insights that will help me take my workout to the next level. But the advertisements are annoying, they interrupt already brief podcasts. "

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity spaCy

**"Makes me so motivated!"** It has a fantastic and enormous variety of sounds and speeds. Very informative and gained some new insights that will help me take my workout to the next level. But the advertisements are annoying, they interrupt already brief podcasts. "

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity **spaCy**

"**Makes me so motivated!** It has a fantastic and enormous variety of **sounds** and **speeds**. Very informative and gained some new insights that will help me take my workout to the next level. But the advertisements are annoying, they interrupt already brief podcasts. "

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity spaCy

"Makes me so **motivated**! It has a fantastic and enormous variety of **sounds** and **speeds**. Very **informative** and gained some **new insights** that will help me take my workout to the next level. But the advertisements are annoying, they interrupt already brief podcasts. "



Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity **spaCy**

"**Makes me so motivated!** It has a fantastic and enormous variety of **sounds** and **speeds**. Very **informative** and gained some **new insights** that will help me take my workout to the next level. But the **advertisements** are annoying, they interrupt already brief podcasts. "

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

[ 1.67, Emotion, Advertisement, Audio ]

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

"**Makes me so motivated!**" [ 1.67, 1.58, Advertisement, Audio ]

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

"**Makes me so motivated!**"

"But the **advertisements** are  
**annoying**, they **interrupt**  
already brief podcasts."

[ 1.67, 1.58, -0.70, Audio ]

Data

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

Review  
Categorization

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

Sentiment Analysis

"**Makes me so motivated!**"

Description  
Vectorization

"But the **advertisements** are  
**annoying**, they **interrupt**  
already brief podcasts."

Prediction

"It has a **fantastic** and  
**enormous** variety of **sounds**  
and **speeds**."

[ 1.67, 1.58, -0.70, 0.90 ]

Data

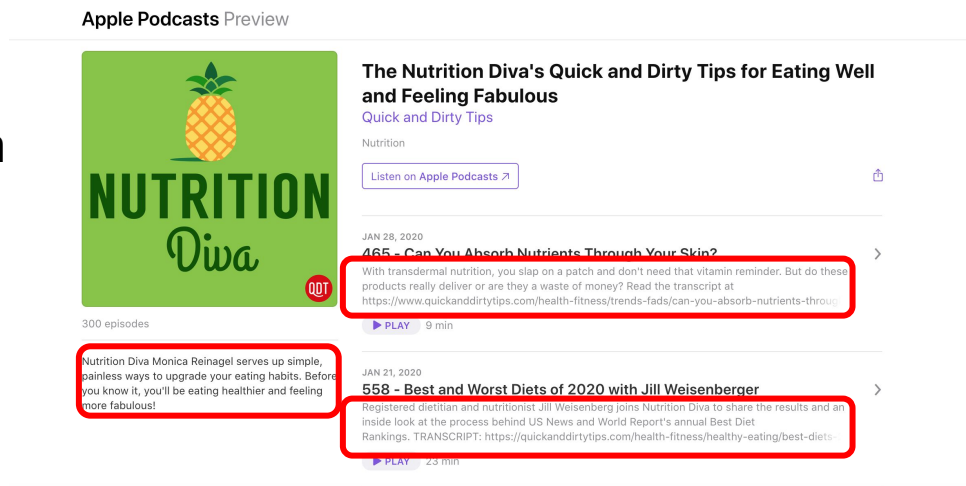
Review  
Categorization

Sentiment Analysis

Vectorization

Prediction

## Description



- TF-IDF (Term Frequency - Inverse Document Frequency) vectorize descriptions
- Concatenate with the sentiment vector

**Description TF-IDF Vector +  
Review Sentiment Vector**

Data

Review  
Categorization

Sentiment Analysis

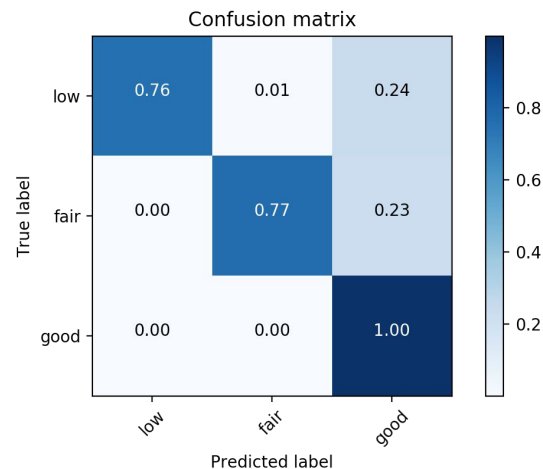
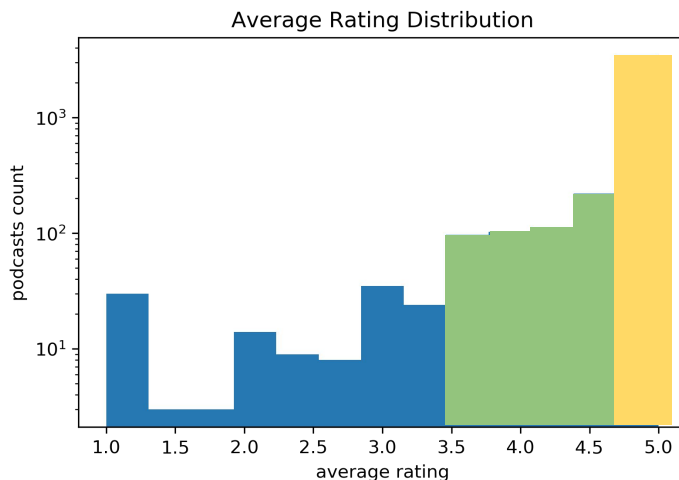
Vectorization

Prediction

**Low:**  $0 < \text{average\_rating} \leq 3.5$  (136 samples)  
**Fair:**  $3.5 < \text{average\_rating} \leq 4.8$  (658 samples)  
**Good:**  $4.8 < \text{average\_rating} \leq 5.0$  (3321 samples)

## Random Forest classification

Train and validate the weighted model on samples  
accuracy **81%**



# How it works?

## Apple Podcasts Preview



### Intensive Care Network Podcasts

Oliver Flower

Medicine

[Listen on Apple Podcasts ↗](#)



JAN 21, 2020

#### Developing EM - Colombia, March 9-11 2020

Oliver Flower talks to Mark Newcomb and Lee Fineberg about the upcoming Developing EM conference in Cartagena, Colombia. March 9-11, 2020. All the details are here: <https://developingem.com>

[▶ PLAY](#) 17 min

300 episodes

Podcasts on topics relevant to intensive care medicine

OCT 3, 2019

#### Top 10 critical care papers of recent times

Top 10 critical care papers of recent times

Dr Paul Young

From CICM Trainee Symposium 2019



Mac

iPad

iPhone

Watch

TV

Music

Support



## Apple Podcasts Preview

### Intensive Care Network Podcasts

#### Customer Reviews



Mankotin, 01/25/2012

##### Very practical

Great podcasts, pretty much the only one on neurocritical care would like to see more of them though, like the format of ICH podcast

More ways to shop: Visit an [Apple Store](#), call 1-800-MY-APPLE, or [find a reseller](#).

Copyright © 2019 Apple Inc. All rights reserved. [Privacy Policy](#) | [Terms of Use](#) | [Sales and Refunds](#) | [Legal](#) | [Site Map](#)

Choose your country or region

## PODCAST *Prophet*

Please Enter Fitness Podcast Link:

Submit

Podcast name: **Intensive Care Network Podcasts**

good

The Podcast is rated as good

Information score: **0.9**

Emotion score: **N/A**

Advertisement score: **N/A**

Audio score: **N/A**

Need more helpful information. 😊

N/A

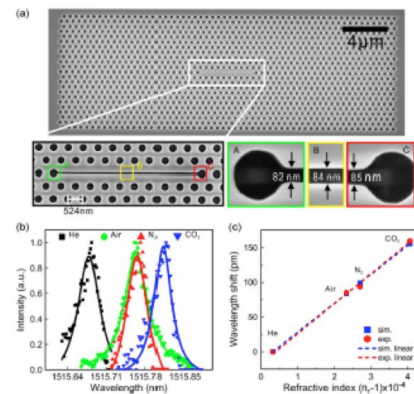
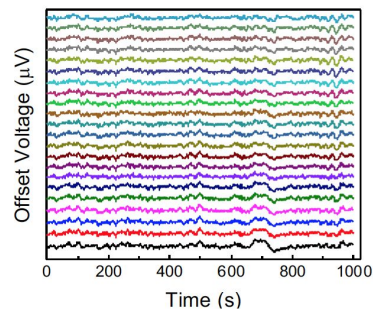
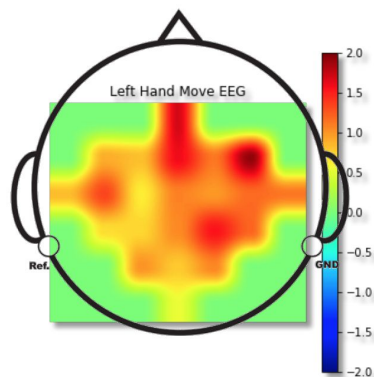
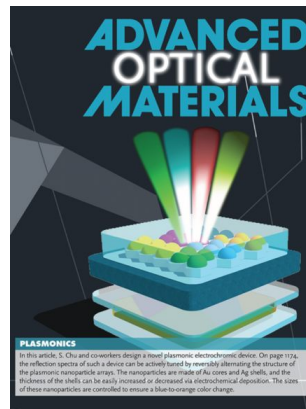
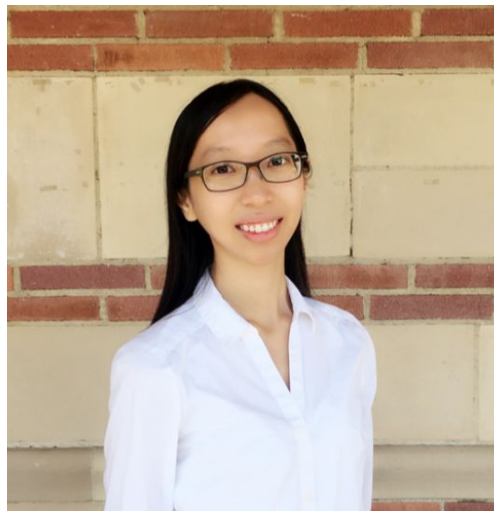
N/A

N/A



# Ziying Feng

PhD, Chemistry



Thank you!

# Web app

<http://datascienceanalysis.xyz/>



**PODCAST** *Prophet*

Please Enter Fitness Podcast Link:

**Submit**

Please input a fitness podcast link. More links at [itunes Preview](#).

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

"Makes me so **motivated**! It has a fantastic and enormous variety of **sounds** and **speeds**. Very **informative** and gained some **new insights** that will help me take my workout to the next level. But the **advertisements** are annoying, they interrupt already brief podcasts. "

**vaderSentiment**

**Review Sentiment Vector**

**[ Information, Emotion, Advertisement, Audio ]**

Data

Review  
Categorization

Sentiment Analysis

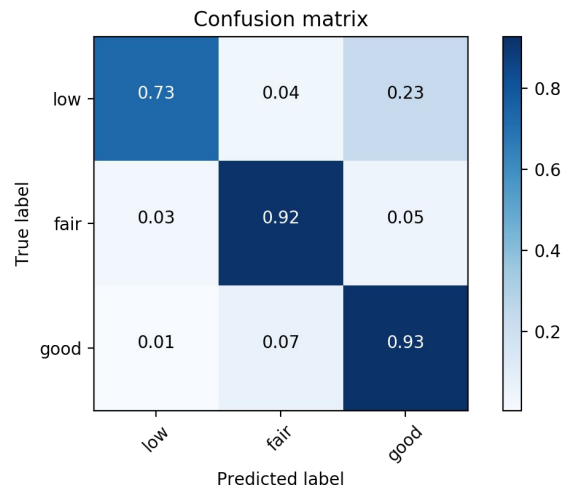
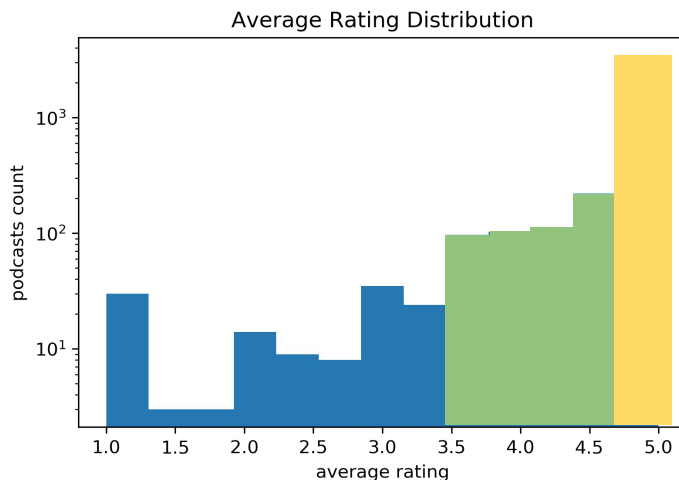
Vectorization

Prediction

## Logistic Regression classification

- Low:**  $0 < \text{average\_rating} \leq 3.5$  (136 samples)  
**Fair:**  $3.5 < \text{average\_rating} \leq 4.8$  (658 samples)  
**Good:**  $4.8 < \text{average\_rating} \leq 5.0$  (3321 samples)

Train and validate the weighted model on samples that have  $\geq 2$  reviews, accuracy **98%**



Data

Review  
Categorization

Sentiment Analysis

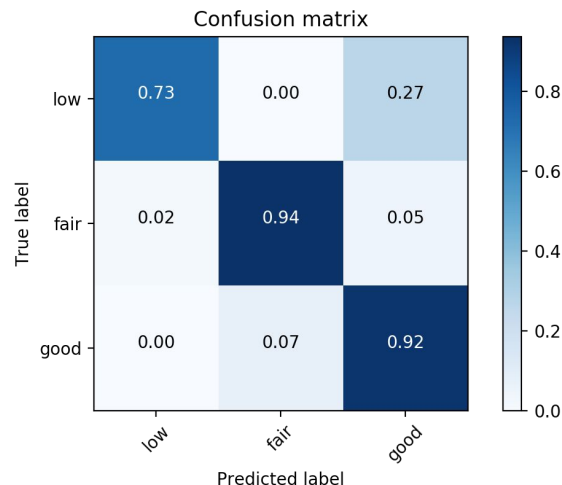
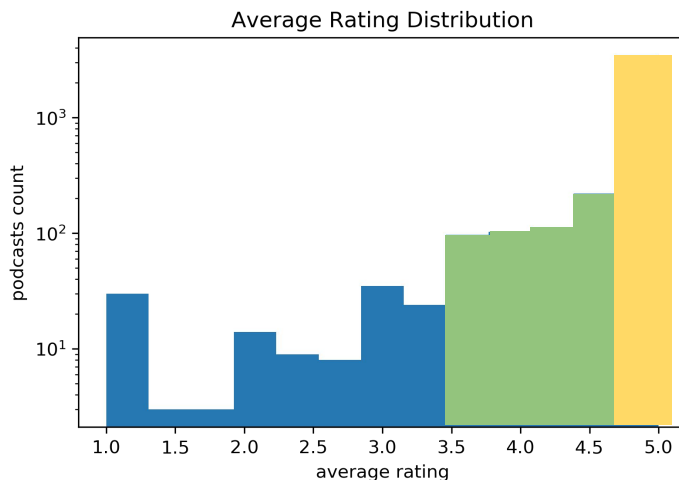
Vectorization

Prediction

## SVM classification

- Low:**  $0 < \text{average\_rating} \leq 3.5$  (136 samples)  
**Fair:**  $3.5 < \text{average\_rating} \leq 4.8$  (658 samples)  
**Good:**  $4.8 < \text{average\_rating} \leq 5.0$  (3321 samples)

Train and validate the weighted model on samples  
that have  $\geq 2$  reviews, accuracy **91%**



Data

Review  
Categorization

Sentiment Analysis

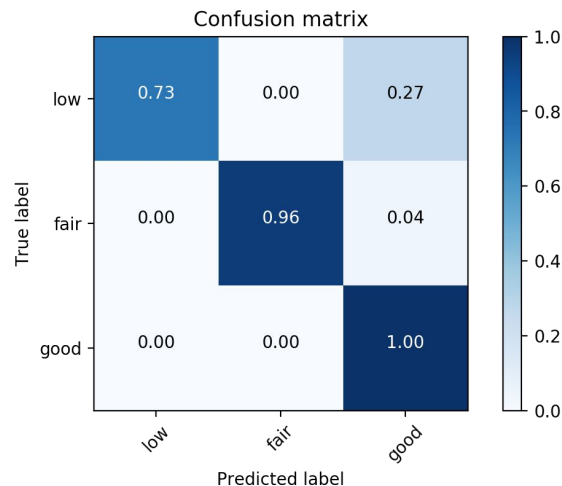
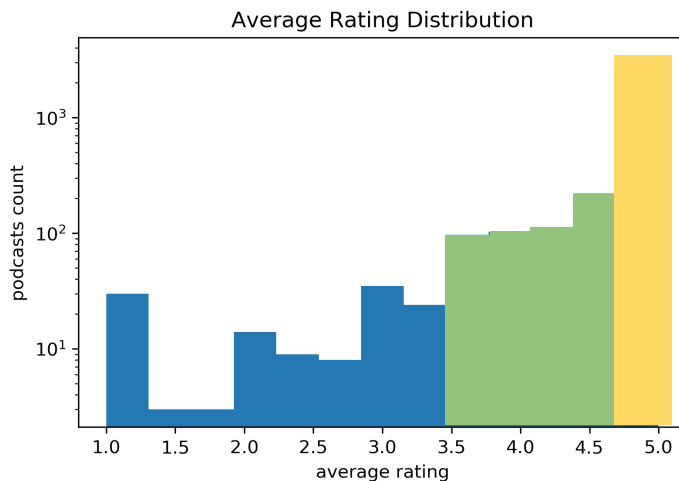
Vectorization

Prediction

## Random Forest classification

- Low:**  $0 < \text{average\_rating} \leq 3.5$  (136 samples)  
**Fair:**  $3.5 < \text{average\_rating} \leq 4.8$  (658 samples)  
**Good:**  $4.8 < \text{average\_rating} \leq 5.0$  (3321 samples)

Train and validate the weighted model on samples that have  $\geq 2$  reviews, accuracy **98%**



Data

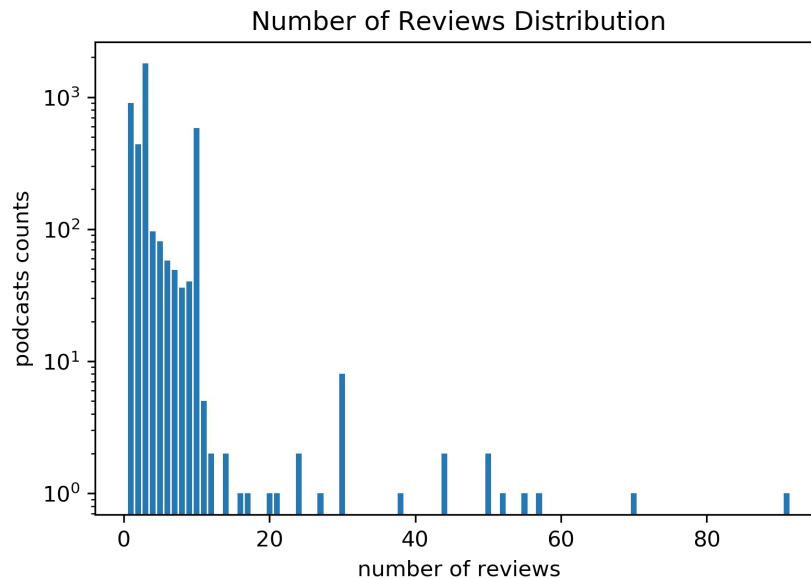
Review  
Categorization

Sentiment Analysis

Vectorization

Prediction

## Logistic Regression vs SVM





Data

Review  
Categorization

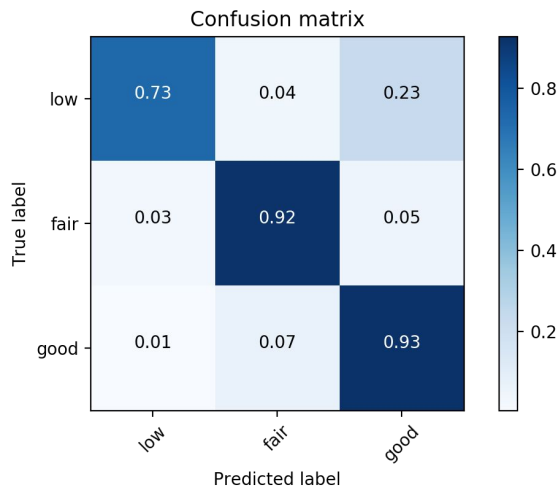
Sentiment Analysis

Vectorization

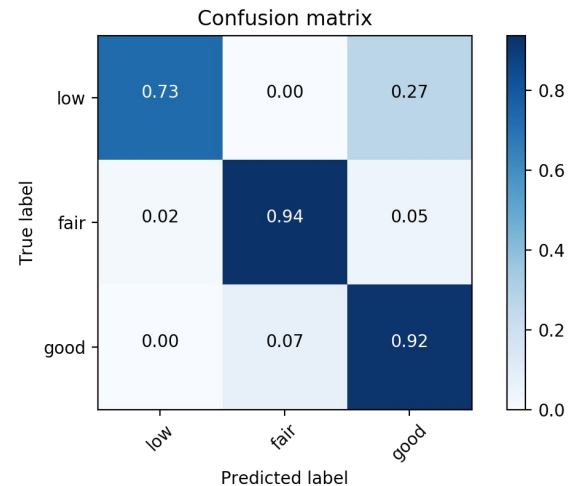
Prediction

## Logistic Regression vs SVM

### Logistic Regression



### SVM



Data

Review  
Categorization

Sentiment Analysis

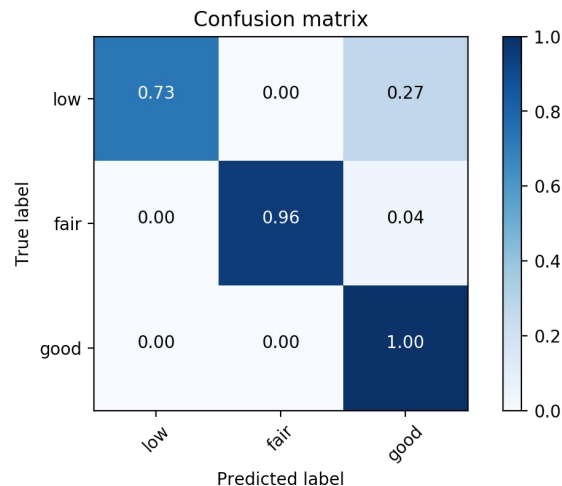
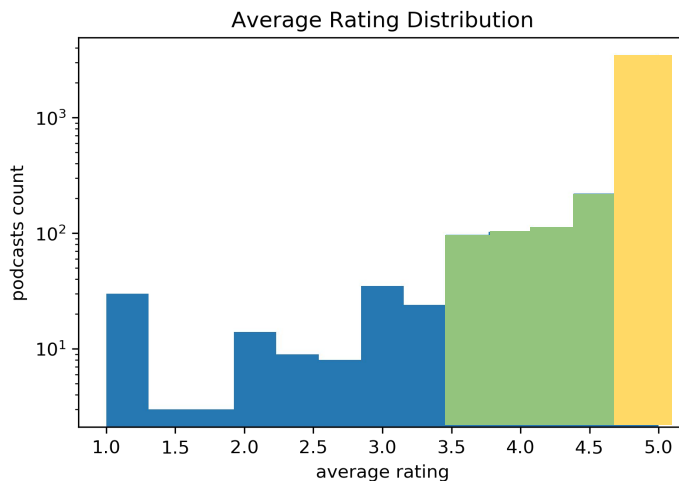
Vectorization

Prediction

## Random Forest classification

- Low:**  $0 < \text{average\_rating} \leq 3.5$  (136 samples)  
**Fair:**  $3.5 < \text{average\_rating} \leq 4.8$  (658 samples)  
**Good:**  $4.8 < \text{average\_rating} \leq 5.0$  (3321 samples)

Train and validate the weighted model on samples that have  $\geq 2$  reviews, accuracy **98%**





# Bag of Words

Doc1: *It was the best of times*

Doc2: *it was the worst of times*

Doc3: *it was the age of wisdom*

Doc4: *it was the age of foolishness*

Unique words:

- “it”
- “was”
- “the”
- “best”
- “of”
- “times”
- “worst”
- “age”
- “wisdom”
- “foolishness”

Doc1: [1, 1, 1, 1, 1, 1, 0, 0, 0, 0]

- “it” = 1
- “was” = 1
- “the” = 1
- “best” = 1
- “of” = 1
- “times” = 1
- “worst” = 0
- “age” = 0
- “wisdom” = 0
- “foolishness” = 0



# TF-IDF

- Term Frequency Inverse Document Frequency
- evaluate how important a word is to a document in a collection or corpus.
- TF = Term Frequency, measures how frequently a term occurs in a document.
- IDF = Inverse Document Frequency, measures how important a term is.

# TF-IDF

TF = Term Frequency

measures how frequently a term occurs in a document.

$TF(t) = (\text{Number of times term } t \text{ appears in a document}) / (\text{Total number of terms in the document})$

5/15

IDF = Inverse Document Frequency

measures how important a term is.

$IDF(t) = \log_e(\text{Total number of documents} / \text{Number of documents with term } t \text{ in it})$

$\log(4/2)$

$TF-IDF = 5/15 * \log(4/2)$

Doc1

Cat ... cat.....  
.....cat.....cat.....  
.....  
.....cat.....

Corpus

Doc1

Cat ... cat.....  
.....cat.....cat.....  
.....  
.....cat.....

Doc2

.....  
.....  
.....  
.....

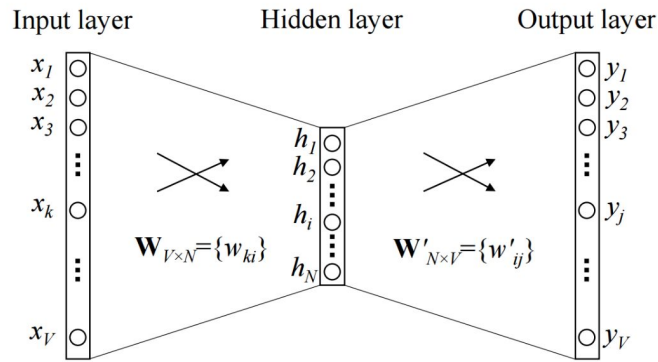
Doc3

.....  
.....  
.....  
.....

Doc4

.....  
.....cat.....  
.....  
.....

# Word2Vec



The drawback of the classical approach is that the context is lost.

Solution: word embedding

It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc.

**Word2vec** is an unsupervised model that learns word embedding from raw text.

**Word2Vec** builds on the fact that words that share similar contexts also share semantic meanings.

**Word2vec** model predicts a word by using its neighbors, by learning dense vectors called embedding.



# Problem



## Apple Podcasts Preview



### Abdômen Absoluto

Abdômen Absoluto Cast

Fitness

[Listen on Apple Podcasts](#)

JAN 25, 2020

#### #1 - Como AFINAR a cintura

Afinal, é possível AFINAR a cintura? Sim e não.

Aqui eu explico melhor como você pode alcançar uma cintura mais fina e harmônica.

Faça parte da Família Abdômen Absoluto: [www.abdomenabsoluto.com.br](http://www.abdomenabsoluto.com.br)

[PLAY](#) 11 min

© Abdômen Absoluto Cast

1 episode

Estratégias e dicas para transformar seu corpo em uma MÁQUINA DE QUEIMAR GORDURA e TRINCAR SUA BARRIGA - sem equipamentos ou dietas malucas e com apenas 7 minutos por dia!

[more](#)

Podcasts > Health & Fitness > Nutrition > Wondery



### The Jillian Michaels Show

Jillian Michaels | Wondery >

[Details](#) [Ratings and Reviews](#) [Related](#)

[Subscribe](#)

★★★★☆ (2,422)

Audio

Nutrition

© Empowered Media LLC



LINKS

[Website](#)  
[Report a Concern](#)

#### From the Provider

Jillian Michaels, America's Health and Wellness show that gives you tools to find health and happiness.

▲ NAME

- 1 Too Famous in Finland!
- 2 Can a Higher IQ Actually Hurt You?
- 3 The Ugly Truth About Diet and Exercise
- 4 How Children Can Change You (with Dr. Will Cole)
- 5 The Vegan Debate Rages On
- 6 This is Not a Resolution!
- 7 Get a Killer Workout at Work?!
- 8 Can Hypnotherapy Really Help You?

Podcasts > Health & Fitness > ACTIONPARK MEDIA



### American Glutton

ACTIONPARK MEDIA >

[Details](#) [Ratings and Reviews](#) [Related](#)

[Subscribe](#)

★★★★★ (1,404)

Audio

Health & Fitness



LINKS

[Website](#)  
[Report a Concern](#)

#### From the Provider

Actor Ethan Suplee was an aspiring plus size model who became a reality star on the show 'The Contender'. On American Glutton, he'll talk to every two decades, taking us on his journey from obesity to everything back plus 50 pounds. Sharing his ever-dogged determination, and the proper way to become a plus size model.

▲ NAME

- 1 Tom Kier, Tactical Training Expert
- 2 Max Discher, Our Everyman
- 3 Dr. Will Cole: Functional Medicine ...
- 4 Kevin Connolly: Actor, Director
- 5 Introducing An American Glutton
- 6 American Glutton Trailer

- Among **10k** fitness and health podcasts, **4115** podcasts have reviews
- How to become an podcast influencer?
- How to get higher popularity?



# Web app - input



PODCAST *Prophet*

Please Enter Fitness Podcast Link:

Submit

Please input a fitness podcast link. More links at [itunes Preview](#).

# Web app - output

Submit

Podcast name: Logical Weight Loss Podcast

good

The Podcast is rated as good

Information score: 4.9

Emotion score: 2.6

Advertisement score: 1.0

Audio score: 1.4

The information is helpful. 👍

The emotion it conveys is quite positive. 👍

The ads are annoying. 😞

Need to adjust the audio quality. 😞

# Data

## Rating level

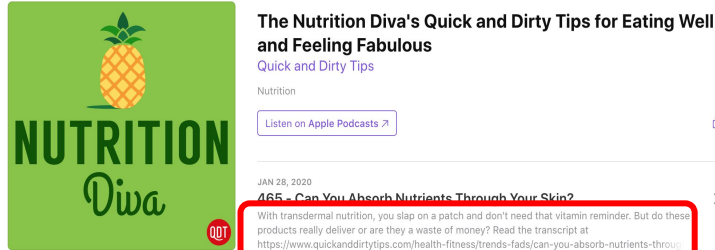
Low:  $0 < \text{average\_rating} \leq 3.5$   
Fair:  $3.5 < \text{average\_rating} \leq 4.9$   
Good:  $4.9 < \text{average\_rating} \leq 5.0$



## Description

## Review

Apple Podcasts Preview



300 episodes

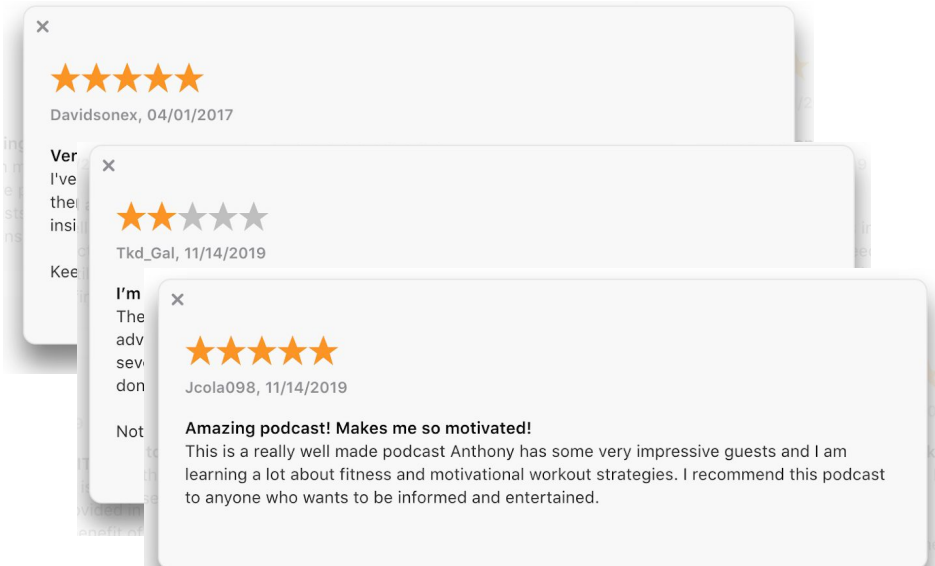
Nutrition Diva Monica Reinagel serves up simple, painless ways to upgrade your eating habits. Before you know it, you'll be eating healthier and feeling more fabulous!

JAN 28, 2020  
**465 - Can You Absorb Nutrients Through Your Skin?**  
With transdermal nutrition, you slap on a patch and don't need that vitamin reminder. But do these products really deliver or are they a waste of money? Read the transcript at <https://www.quickanddirtytips.com/health-fitness/trends-fads/can-you-absorb-nutrients-through-your-skin/>

PLAY 9 min

JAN 21, 2020  
**558 - Best and Worst Diets of 2020 with Jill Weisenberger**  
Registered dietitian and nutritionist Jill Weisenberger joins Nutrition Diva to share the results and an inside look at the process behind US News and World Report's annual Best Diet Rankings. TRANSCRIPT: <https://quickanddirtytips.com/health-fitness/healthy-eating/best-diets-of-2020/>

PLAY 23 min



Data

## Review Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

### Information

knowledge  
data  
news  
intelligence  
insight  
.....

### Emotion

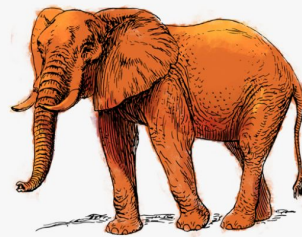
inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

### Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

### Audio

audio  
sound  
voice  
acoustic  
frequency  
.....



### Related Words

Type a word or short phrase...



examples: winter, understanding, cloud

This tool helps you find words that are related to a specific word or phrase.  
Also check out [ReverseDictionary.org](https://www.reverse-dictionary.org/) and [DescribingWords.io](https://describingwords.io/).

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

Information

knowledge  
data  
news  
intelligence  
insight  
.....

Emotion

inspiring  
feeling  
delight  
disappoint  
unhappy  
.....

Advertisement

advertisement  
promotion  
sponsor  
product  
brochure  
.....

Audio

audio  
sound  
voice  
acoustic  
frequency  
.....

Cosine Similarity **spaCy**

× **Information**



Davidsonex, 04/01/2017

**Very informative and fun too!**

I've been training at the Burbank location since it opened, so I thought I knew everything there was to know - but after listening to a few episodes I have really gained some new insights that will help me take my workout to the next level.

Keep up the good work - really loving it!

× **Advertisement**



Tkd\_Gal, 11/14/2019

**I'm going to pass**

The information is interesting. While I understand the need for revenue through advertisements, it is annoying that 1) they interrupt already brief podcasts and 2) of the several I listened to, it is odd that none of her advertisers were healthy food makers. I really don't care that Monica bought a Burberry sweater through one of the advertisers!

Not worth subscribing to at all.

× **Emotion**



Jcola098, 11/14/2019

**Amazing podcast! Makes me so motivated!**

This is a really well made podcast Anthony has some very impressive guests and I am learning a lot about fitness and motivational workout strategies. I recommend this podcast to anyone who wants to be informed and entertained.

× **Audio**



SaEmGaLL, 04/05/2019

**Use it and recommend it weekly**

Podunner has a fantastic and enormous variety of sounds and speeds, and is so easy to use. Every week someone around me says they need walking or running music, and I enthusiastically recommend Podrunner every time. Hands down. Thank you, Steve Boyett!

Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

X

★★★★★

Davidsonex, 04/01/2017

**Very informative and fun too!**

I've been training at the Burbank location since it opened, so I thought I knew everything there was to know - but after listening to a few episodes I have really gained some new insights that will help me take my workout to the next level.

Keep up the good work - really loving it!

X

★★★★★

Jcola098, 11/14/2019

**Amazing podcast! Makes me so motivated!**

This is a really well made podcast Anthony has some very impressive guests and I am learning a lot about fitness and motivational workout strategies. I recommend this podcast to anyone who wants to be informed and entertained.

X

★★★☆☆

Tkd\_Gai, 11/14/2019

**I'm going to pass**

The information is interesting. While I understand the need for revenue through advertisements, it is annoying that 1) they interrupt already brief podcasts and 2) of the several I listened to, it is odd that none of her advertisers were healthy food makers. I really don't care that Monica bought a Burberry sweater through one of the advertisers!

Not worth subscribing to at all.

X

★★★★★

SaEmGaLL, 04/05/2019

**Use it and recommend it weekly**

Podrunner has a fantastic and enormous variety of sounds and speeds, and is so easy to use. Every week someone around me says they need walking or running music, and I enthusiastically recommend Podrunner every time. Hands down. Thank you, Steve Boyett!

[ 1.67, 0, 0, 0 ]

[ 1.67, 1.58, 0, 0 ]

[ 1.67, 1.58, -0.70, 0 ]

[ 1.67, 1.58, -0.70, 0.90 ]

Data

**vaderSentiment**

**[ Information, Emotion, Advertisement, Audio ]**

Review  
Categorization

"Very informative. Gained some new insights that will help me take my workout to the next level."

**[ 1.67, 0, 0, 0 ]**

Sentiment Analysis

"Amazing podcast! Makes me so motivated!"

**[ 1.67, 1.58, 0, 0 ]**

Description  
Vectorization

"The advertisements are annoying, they interrupt already brief podcasts"

**[ 1.67, 1.58, -0.70, 0 ]**

Prediction

"Podunner has a fantastic and enormous variety of sounds and speeds"

**[ 1.67, 1.58, -0.70, 0.90 ]**

Data

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

Review  
Categorization

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

[ 1.67, 0, 0, 0 ]

Sentiment Analysis

"**Makes me so motivated!**"

[ 1.67, 1.58, 0, 0 ]

Description  
Vectorization

"But the **advertisements** are  
**annoying**, they **interrupt**  
already brief podcasts."

[ 1.67, 1.58, -0.70, 0 ]

Prediction

"It has a **fantastic** and  
**enormous** variety of **sounds**  
and **speeds**."

[ 1.67, 1.58, -0.70, 0.90 ]



Data

Review  
Categorization

Sentiment Analysis

Description  
Vectorization

Prediction

vaderSentiment

[ Information, Emotion, Advertisement, Audio ]

"**Very informative** and  
gained some **new insights**  
that will help me take my  
workout to the next level."

[ 1.67, Emotion, Advertisement, Audio ]

"**Makes me so motivated!**"

[ 1.67, 1.58, Advertisement, Audio ]

"But the **advertisements** are  
**annoying**, they **interrupt**  
already brief podcasts."

[ 1.67, 1.58, -0.70, Audio ]

"It has a **fantastic** and  
**enormous** variety of **sounds**  
and **speeds**."

[ 1.67, 1.58, -0.70, 0.90 ]

Data

Review  
Categorization

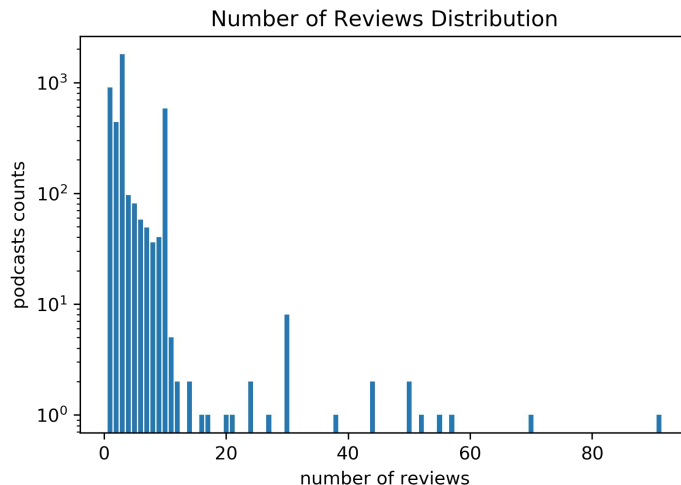
Sentiment Analysis

Vectorization

Prediction

## SVM classification

- Train the **weighted model** on **2771** samples having  $> 2$  reviews
- Validate on the **1334** samples having  $\leq 2$  reviews
- Accuracy **82%**



Data

Review  
Categorization

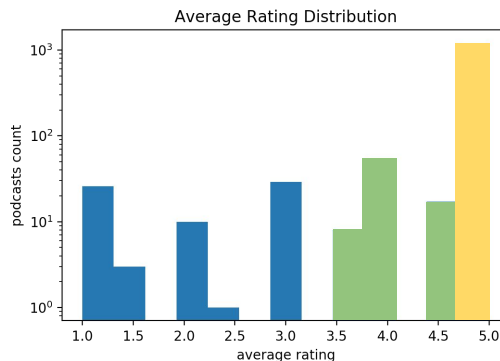
Sentiment Analysis

Vectorization

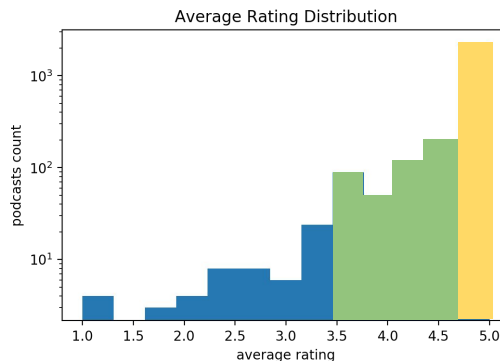
Prediction

## SVM classification

- Train on samples that have  $> 2$  reviews
- Validate on samples that have  $\leq 2$  reviews



Low rating level mainly  
distributes within the samples  
with  $\leq 2$  reviews



Data

Review  
Categorization

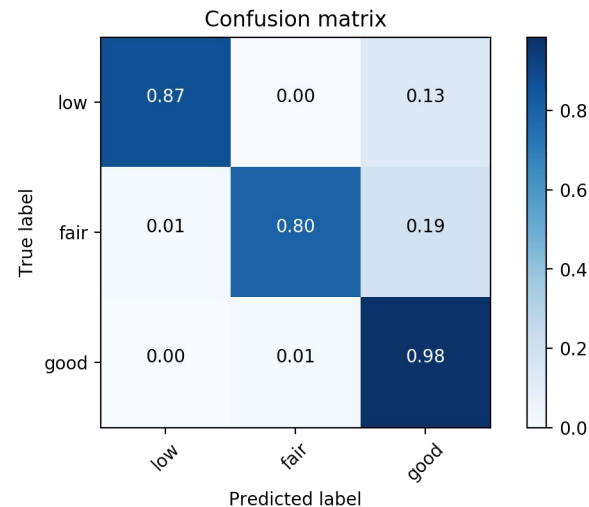
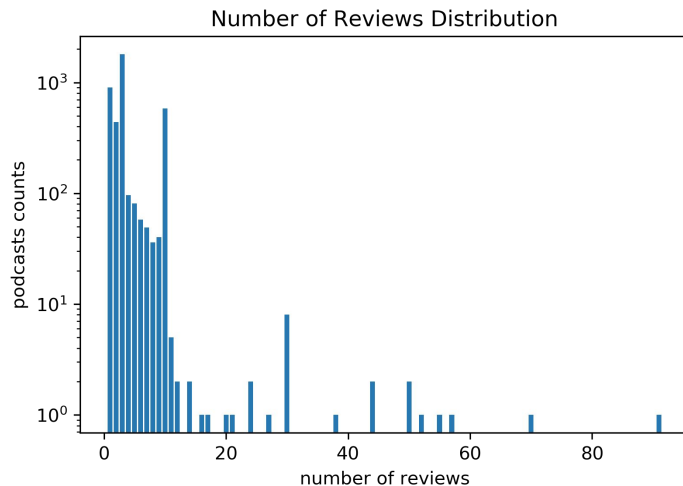
Sentiment Analysis

Vectorization

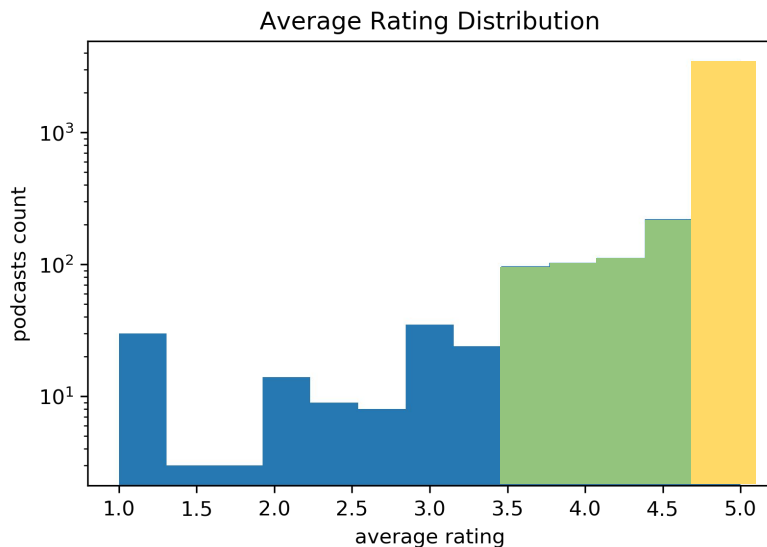
Prediction

## SVM classification

- Validate model on the samples having  $\leq 2$  reviews



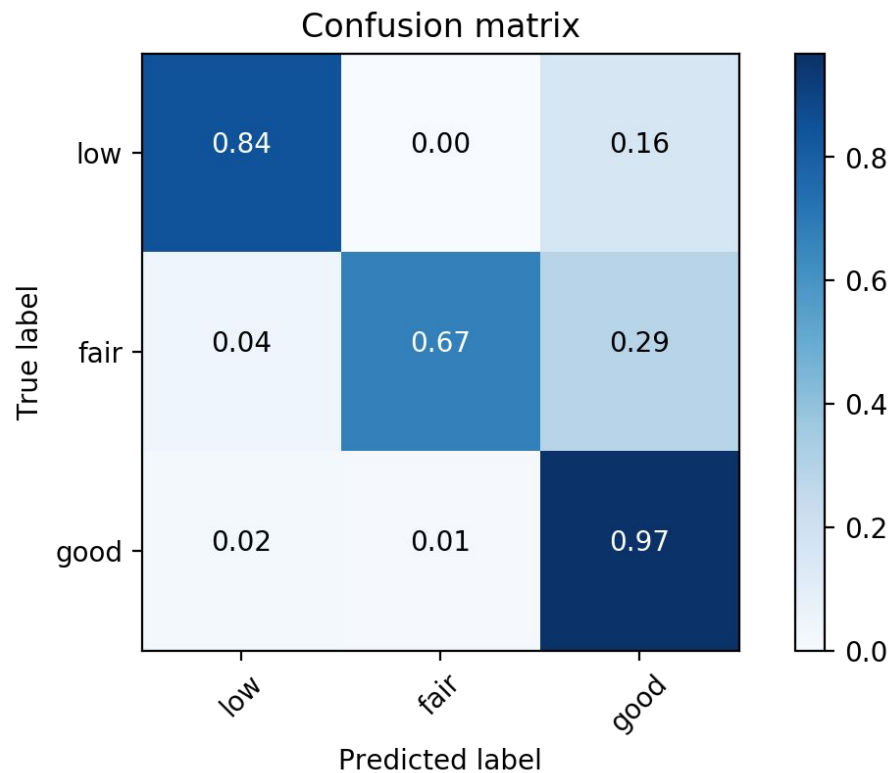
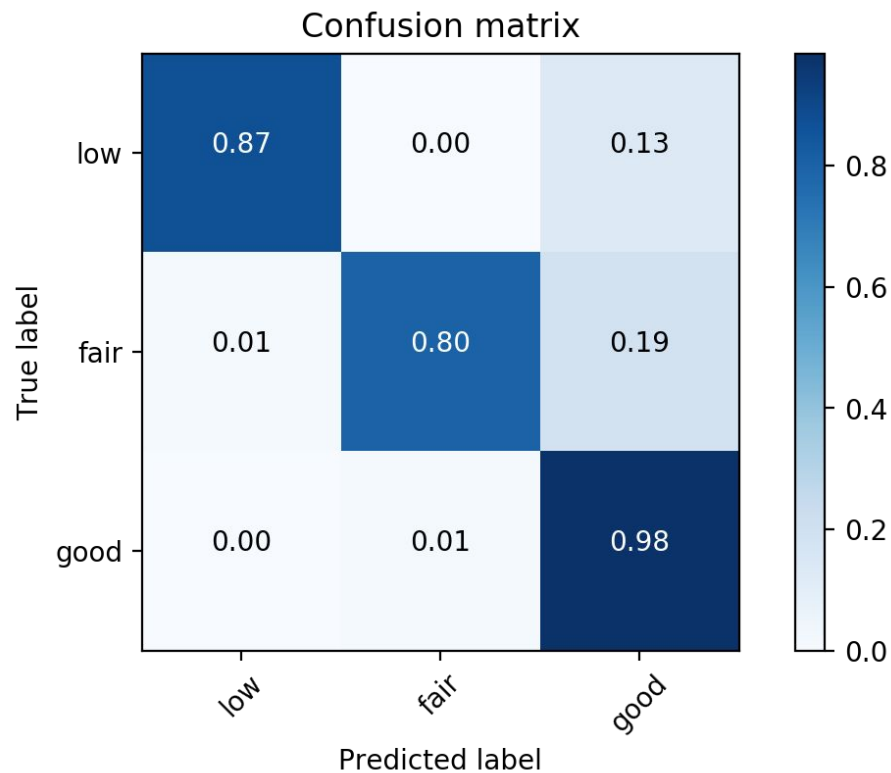
# Data



## Rating level

- Low: **0** < average\_rating <= **3.5** (136 samples)  
Fair: **3.5** < average\_rating <= **4.8** (658 samples)  
Good: **4.8** < average\_rating <= **5.0** (3321 samples)

# Train on train set validate on less2



Train on over2, validate on less2

Train on train set, validate on test less2

Train on train set, validate on less2