

Welcome to Julie's Portfolio

01 About Me

Self-introduction

02 Professional Experience

Background, experience, skills

03 Deep Learning Projects

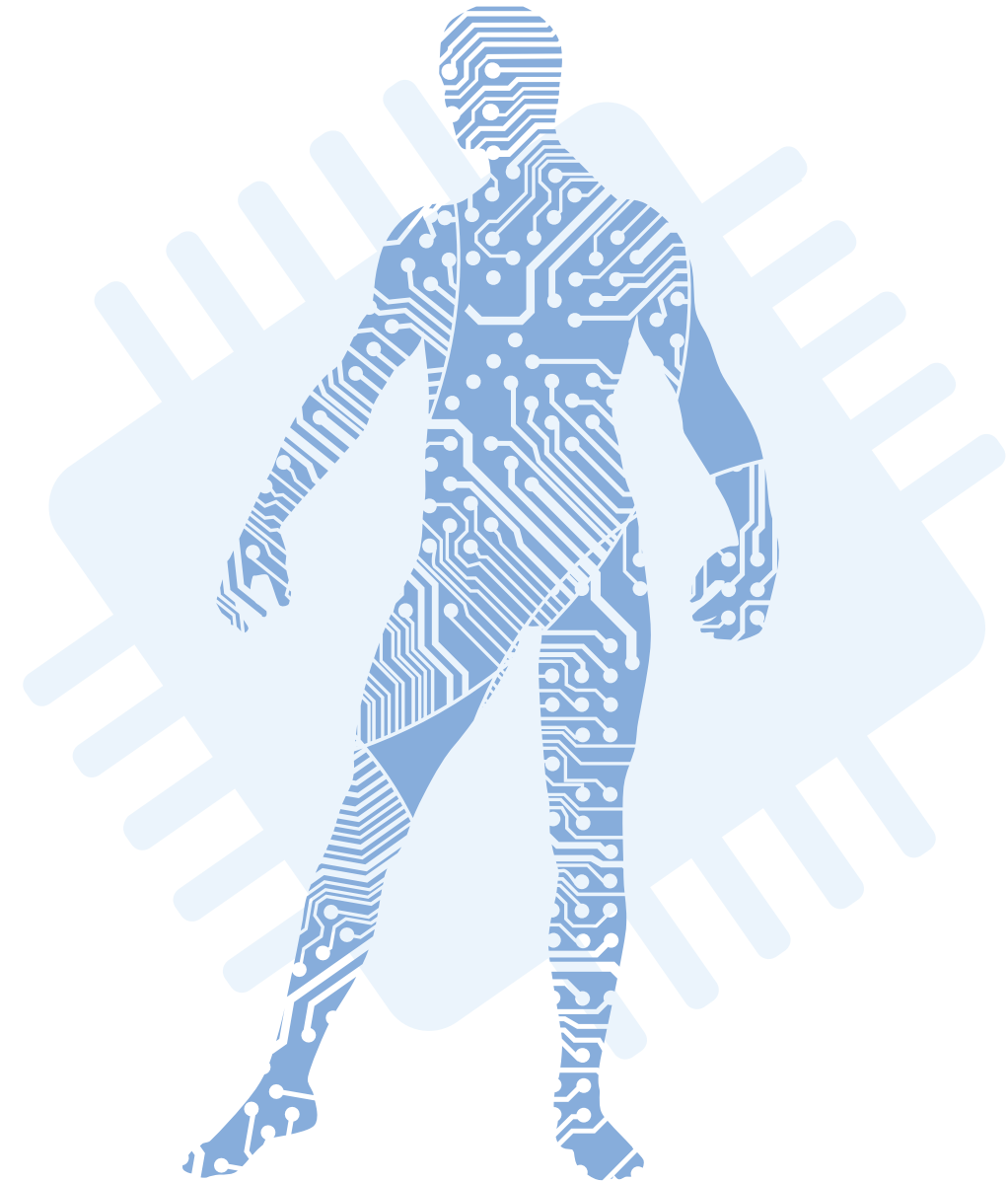
Presentation of deep learning works

04 Machine Learning Projects

Presentation of machine learning works

05 Summary

Portfolio summary





About Me

Julie Yan

Machine Learning Engineer
QA Engineer
Sports Lover

In 2015, I joined MSC software as a QA engineer, when I graduate from University of Southern California, and got my master degree. In 2016, msc former CEO, Dominic Gallelo encouraged everybody to study machine learning, which discipline gets more and more popular and powerful. Online courses are provided to all employees, and that was the start point of my journey in artificial intelligence.

Over the past three years, I have dedicated a considerable amount of time and effort to machine learning. Work hard to equip and prepare myself to be a skilled machine learning engineer. I have completed a number of courses in machine learning, deep learning, data science, statistics, probability, computer science, data structure and algorithm; as well as challenging projects on computer vision and natural language processing.

Time fly, hopefully my career in machine learning can take off too.

Experience

2013-2015

University of Southern California

Mechanical Engineering
Master of Science

2016-future

Machine Learning

Start journey of artificial intelligence,
machine learning, deep learning and
data science

2009-2013

Beihang University

Aerospace Engineering
Bachelor of Science

2015-current

MSC Software Corporation

QA Engineer

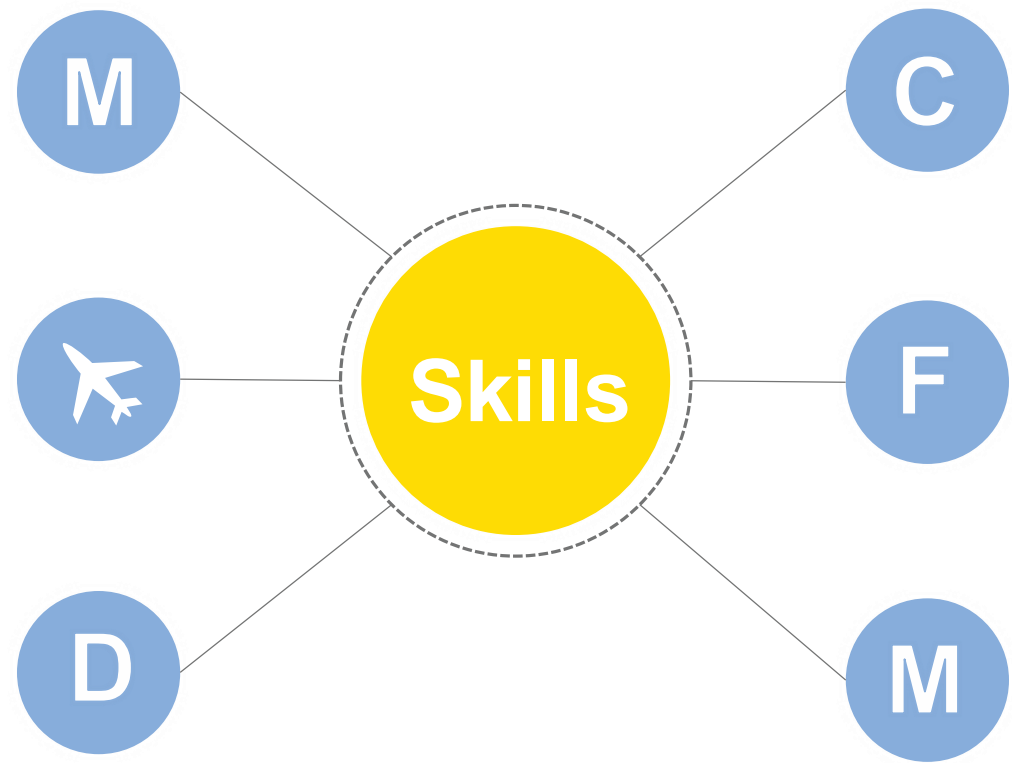


Machine Learning

- Linear Regression
- Logistic Regression
- Decision Tree
- KNN
- K-Means
- KD-tree
- Support Vector Machines
- Naïve Bayes
- Random Forest
- PCA
- Adaboost
- Gradient Boost
- XGBoost

Deep Learning

- CNN
- R-CNN
- Fast R-CNN
- Faster R-CNN
- Mask R-CNN
- YOLO
- VGG
- ResNet
- Unet
- Deconvolution Net
- RNN
- LSTM
- GRU
- Attention
- Bi-directional



Computer Science

- Data Structure
- Data Algorithm
- Python
- Windows / Linux

Framework

- Tensorflow
- Keras
- Scikit-learn
- Nltk
- Numpy
- Pandas
- Scipy
- Matplotlib
- Beautiful soup
- cv2

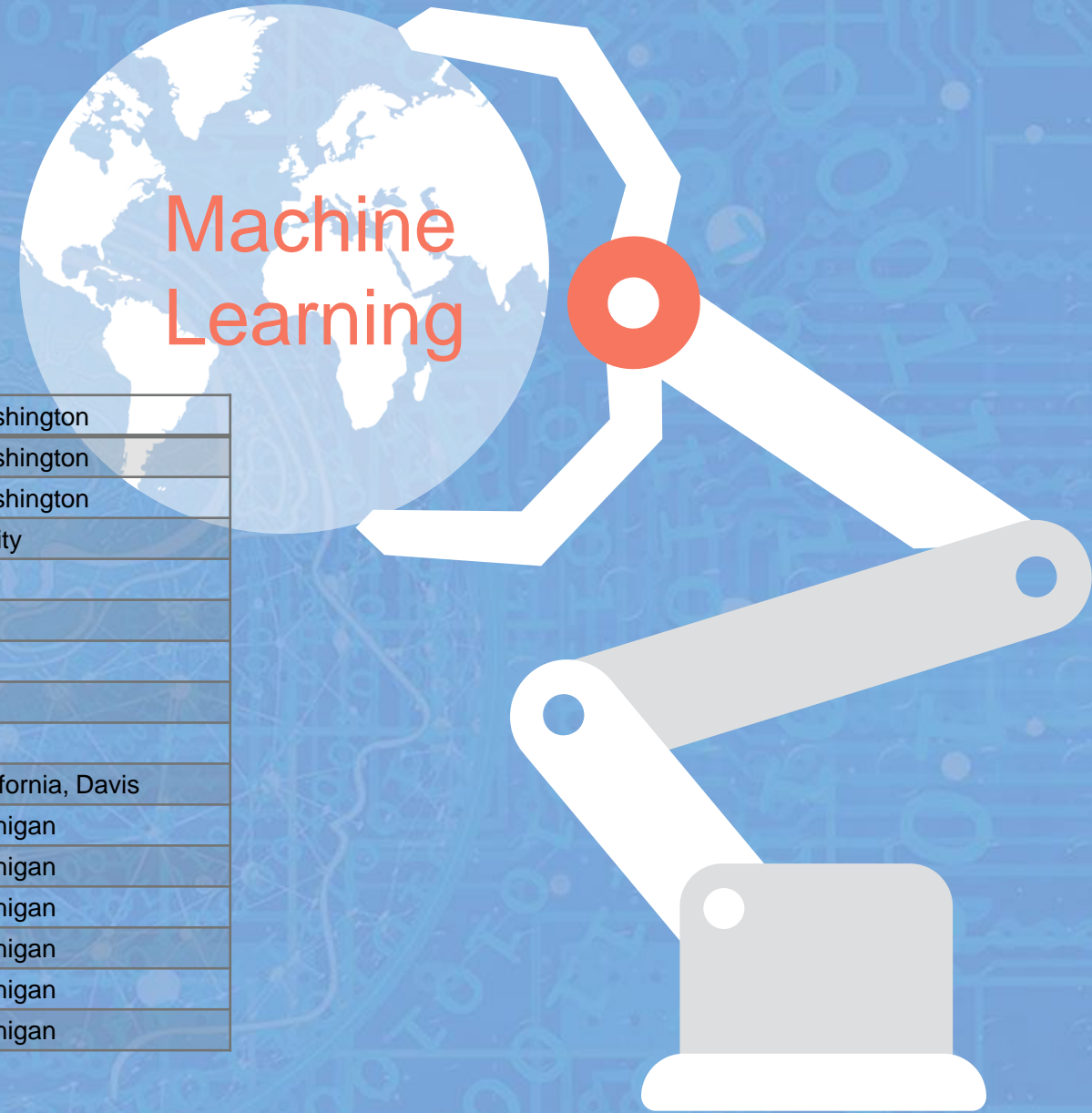
Mathematics

- Statistics
- Probability
- Linear Algebra
- Applied Mathematics
- Engineering Mathematics

Certificates

In the past three years, I have dedicated a considerable time and efforts on machine learning and deep learning. Here is a list of certificates I achieved for data science and machine learning.

• Machine Learning Classification	University of Washington
• Machine Learning Regression	University of Washington
• Machine Learning Clustering & Retrieval	University of Washington
• Machine Learning	Stanford University
• Neural Networks and Deep Learning	Deeplearning.AI
• Structuring Machine Learning Projects	Deeplearning.AI
• Convolutional Neural Network CNN	Deeplearning.AI
• Recurrent Neural Network RNN	Deeplearning.AI
• Hyper parameter tuning, Regularization and Optimization	Deeplearning.AI
• SQL for Data Science	University of California, Davis
• Python Data Structures	University of Michigan
• Using Python to Access Web Data	University of Michigan
• Using Databases with Python	University of Michigan
• Retrieving, Processing, and Visualizing Data with Python	University of Michigan
• Data Science in Python	University of Michigan
• Applied Plotting, Charting & Data Representation in Python	University of Michigan





Deep Learning Projects

DNN CNN RNN Project Presentation



Segmentation Project

The task is to do both categorization and segmentation of rich and complete apparel attributes, an important step toward real-world applications.

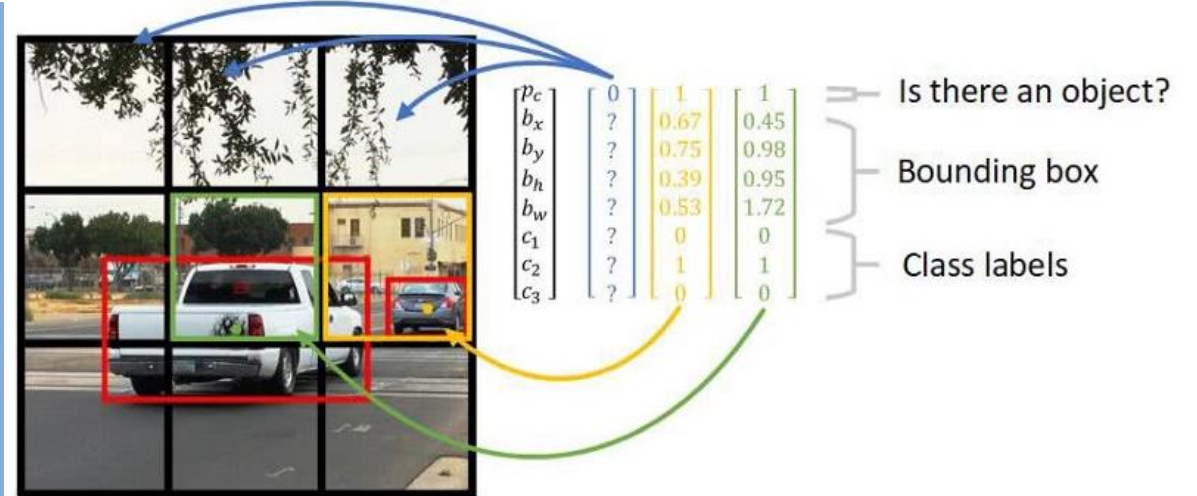
- Project: Kaggle - iMaterialist (Fashion) 2019 at FGVC6, <https://www.kaggle.com/c/imaterialist-fashion-2019-FGVC6>
- Model Structure: **Mask RCNN**
- Code : <https://github.com/JialieY/Kaggle-iMaterialist>

Object Detection Project

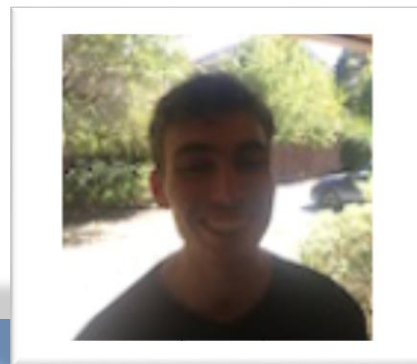
The task is to detect objects like traffic lights, truck, car, and pedestrian in the image and videos.

- Project: Autonomous driving Car detection
- Model Structure: **Yolo model**
- Code :
https://github.com/JialieY/dl_ng/blob/master/Course4/Week3/Autonomous%20driving%20application%20-%20Car%20detection%20-%20v1.ipynb

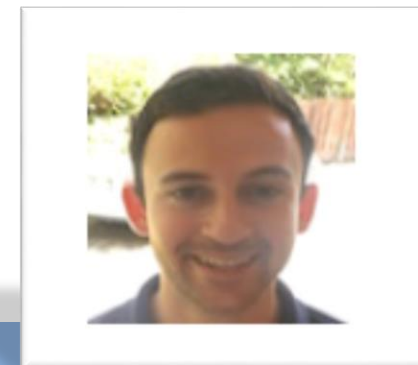
How Does YOLO Work?



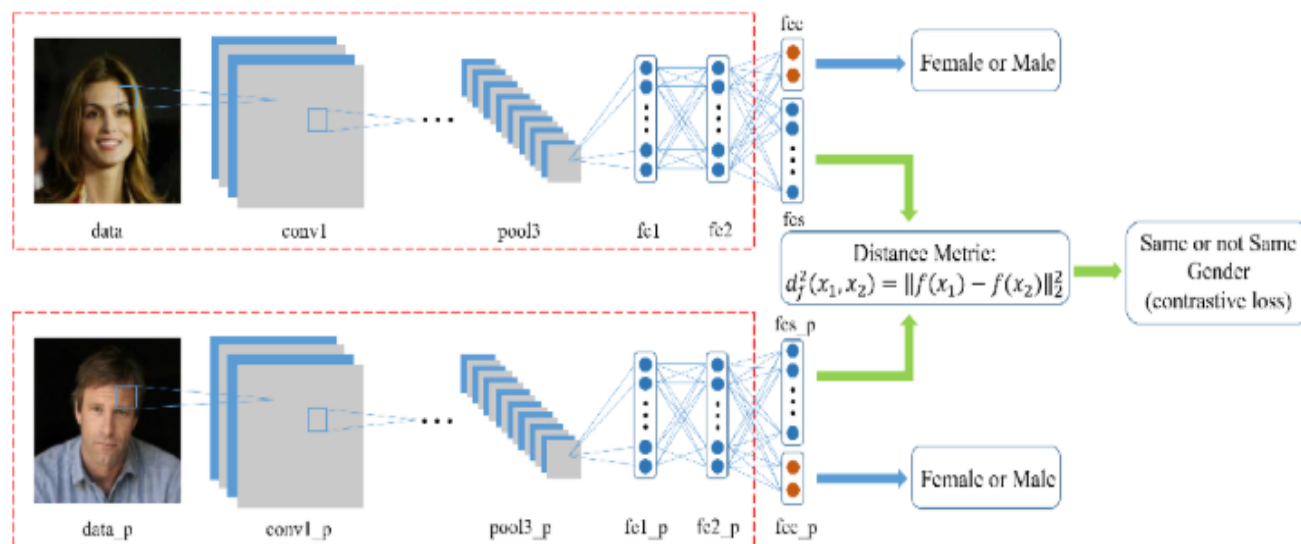
Face Recognition & Verification Project



It's not Kian, please go away
(0.86224014, False)



It's younes, welcome home!
(0.65939283, True)



The task is to recognize and verify people with existing images in the database.

- Project: Face Recognition & Verification
- Model Structure: **FaceNet**, **DeepFace**, **Siamese network**
- Code :

https://github.com/JialieY/dl_ng/blob/master/Course4/Week4/Face/Face%20Recognition%20for%20the%20Happy%20House%20-%20v2.ipynb

Neural Style Transfer Project

The task is to generate a new image with given content and given art style.

- Project: Art generation
- Model Structure: VGG16, VGG19
- Code :
https://github.com/JialieY/dl_ng/blob/master/Course4/Week4/Neural%20Style%20Transfer/Art%20Generation%20with%20Neural%20Style%20Transfer%20-%20v1.ipynb

content image



louvre museum

style image



impressionist style painting

+

||

generated image

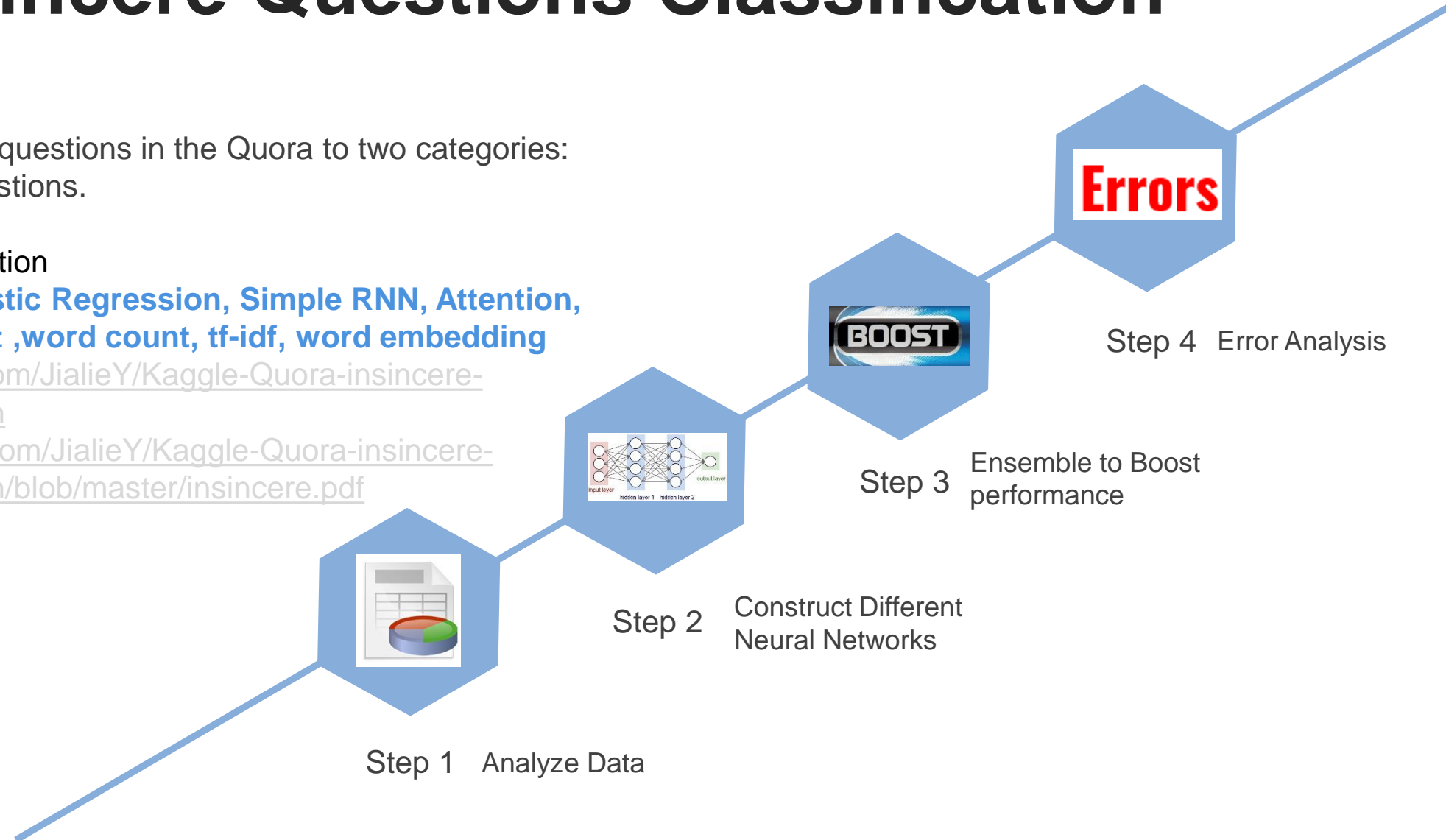


louvre painting
with impressionist style

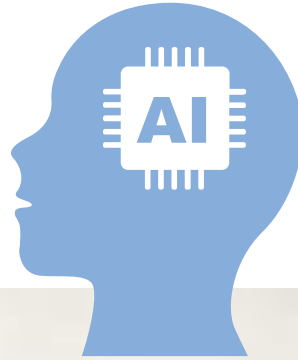
Quora Insincere Questions Classification

The task is to classify the questions in the Quora to two categories: sincere and insincere questions.

- Project: Text Classification
- Model Structure: **Logistic Regression, Simple RNN, Attention, LSTM, GRU, XGboost, word count, tf-idf, word embedding**
- Code : <https://github.com/JialieY/Kaggle-Quora-insincere-questions-classification>
- Report: <https://github.com/JialieY/Kaggle-Quora-insincere-questions-classification/blob/master/insincere.pdf>



I can write poem.



*Dark night,
love brings light to my mind,
when boomers post in the restating ming,
all love hear a thine say which thy tour
thy to the sweat of tonged is hieg my beariage,
my let crefinds in the my polinamed night.
to youth me tame so is of to sone are rode.
and thou lin of eye's sake my sweet fape spoin:
where on shied,
i farnipress lite not miseours an teplife,
for ming ventao en beauty ming with,
tame far that to said*



Sequence Generation Project

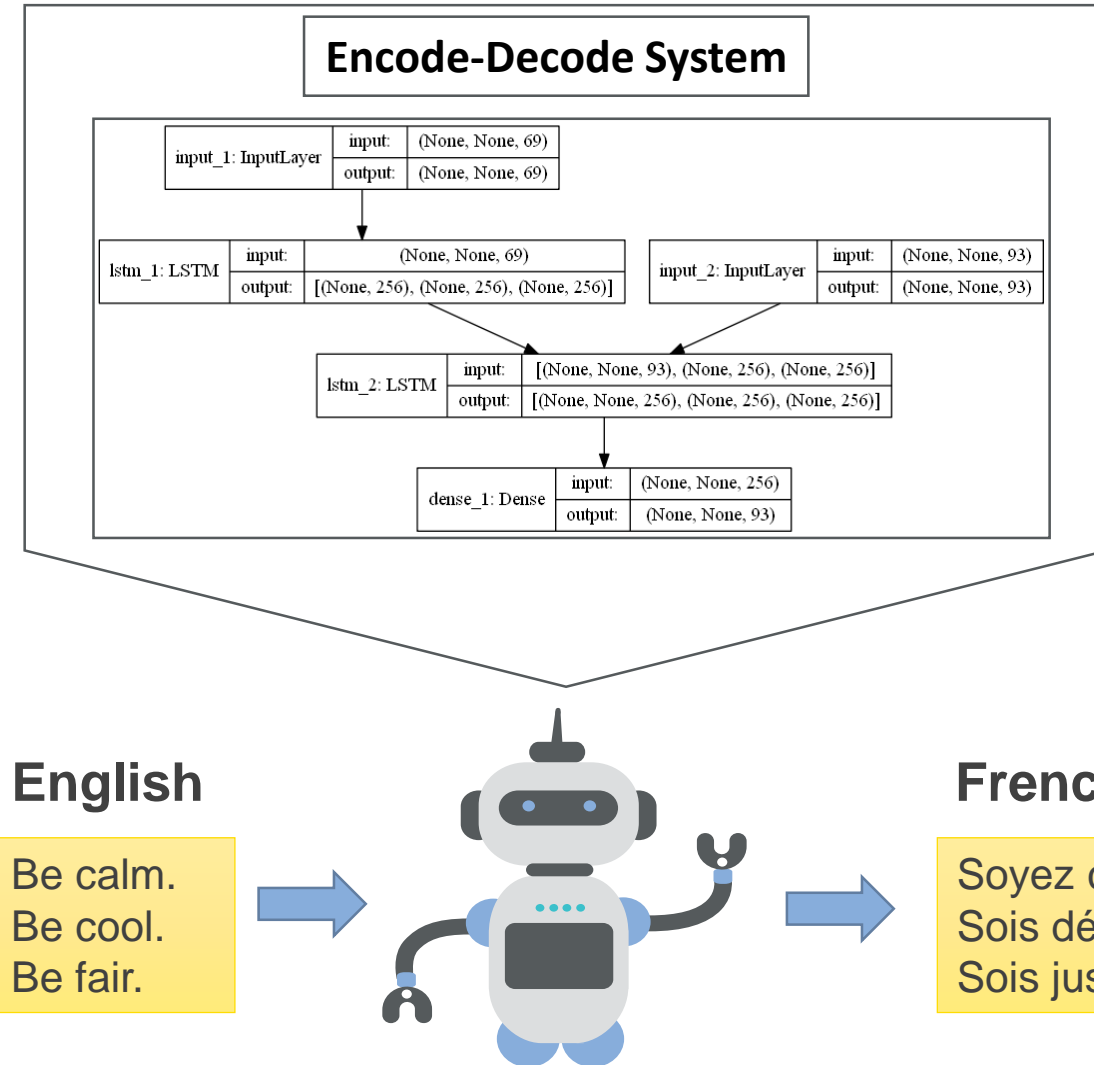
The technique is to learn the given pattern, and generate the similar output.

- Project: Name generation, Article generation, Music generation
- Model Structure: **RNN, LSTM, Attention**
- Code :
https://github.com/JialieY/dl_ng/blob/master/Course5/Week1/Jazz
https://github.com/JialieY/dl_ng/tree/master/Course5/Week1/Dinosaur

Translator Project

The task is to translate English sentence to French sentence.

- Project: English – French Translator
- Model Structure: **LSTM, Encode-Decode system**





Machine Learning Projects

Supervised & unsupervised learning
Projects presentation

Amazon Review Classification

The task is to predict whether the sentiments about a product are positive or negative.

- Project: Amazon product reviews classification
- Data input: Reviews
- Model Structure: Linear regression, Logistic regression, bag of words, tf-idf, L2 regularization, L1 regularization, Stochastic Gradient Descent

Problem

3.7 ★★★★★ 3 reviews

Sort by: Most relevant

Rebecca Ponders
8 reviews

★★★★★ 2 years ago

Horrible customer service. Receptionist is extremely rude. I have called regarding a roofing report weekly for over a month and never received the report. I asked to hold to speak with a manager or the owner and was told by the receptionist that she didn't control have control over their availability and that no one could help me right now, and promptly hung up on me. Truly horrific customer service.

Like

Response from the owner 11 months ago

Ms. Ponders, I want to start off by sincerely apologizing for your negative experience with our company back in 2016. We strive to present top-notch customer service in all department. Our company has undergone numerous changes and have since then improved our customer service policies. It is reviews such as yours that provides our company the ability to learn from our mistakes in hopes of creating a better customer experience in the future. Thank you and I hope you will grant us the opportunity to improve your image of us.

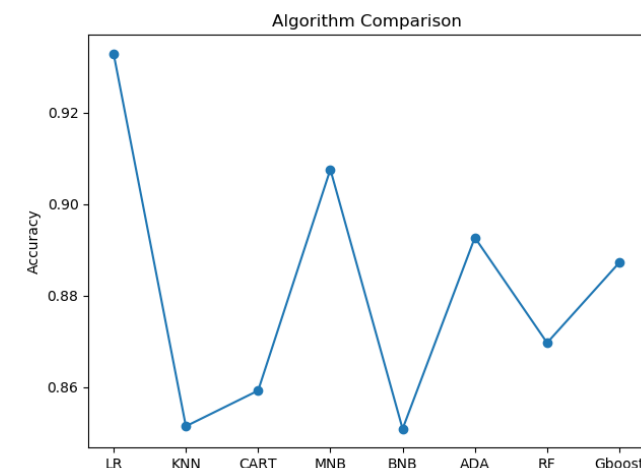
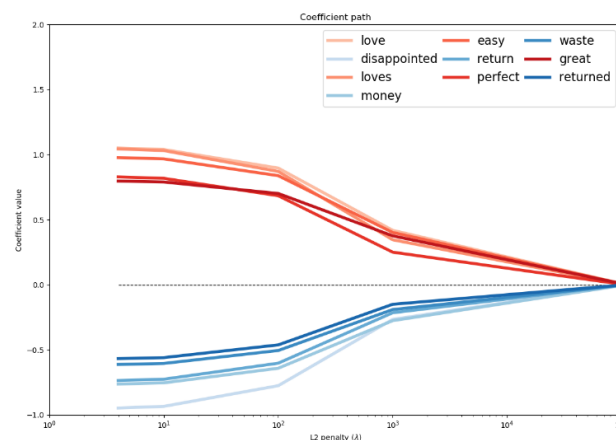
Find out it is



Input

review	rating	sentiment
All of my kids have cried non-stop when I tried to ween them off their pacifier, until I found Thumbuddy To Love	5	1
We wanted to get something to keep track of our child's milestones and this is a cute option. There aren't many c	5	1
My daughter had her 1st baby over a year ago. She did receive and fill up a First Year Calendar. When her son was	5	1
One of baby's first and favorite books, and it is washable! I gave 1 less star than perfect because I'd like to see so	4	1
Very cute interactive book! My son loves this book! The bright colorful illustrations make this great for babies &	5	1
Beautiful book, I love it to record cherished times in my great granddaughters life with the beautiful pastel pink	5	1
Try this out for a spring project !Easy ,fun and affordable wall decals ...Fine quality and brightens up any room.. 5	5	1
very nice Divine Mercy Pendant of Jesus now on chain around my neck. Love It! Divine Mercy represents Jesusw	5	1
We bought the pins as my 6 year old Autistic son was able to open or break open normal safety pins we needed f	4	1
It has been many years since we needed diaper pins, but I do like to pin my socks together through the wash and	5	1
We found this book at a rummage sale and found it to be so useful, especially when several people are helping v	5	1

Algorithm



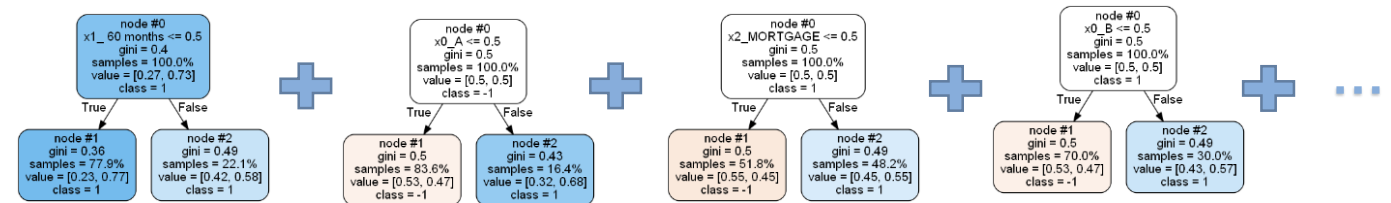
Loan Classification

The task is to predict whether a loan is safe or not.

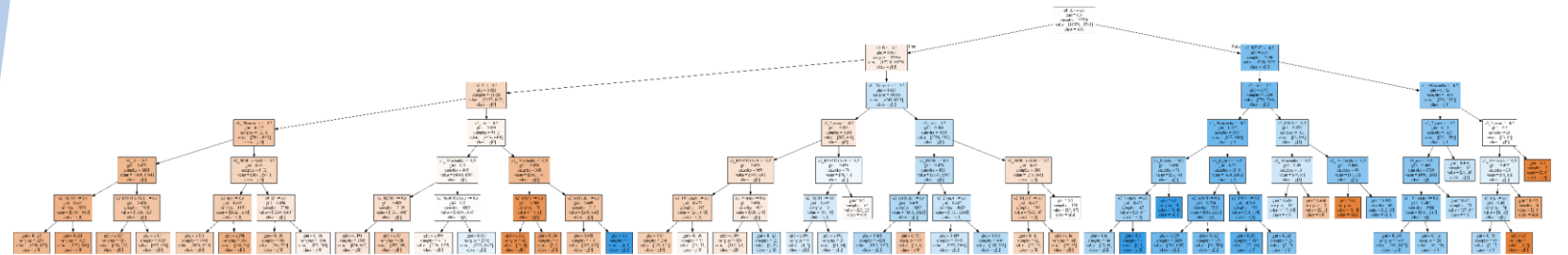
- Project: Loan data classification
- Data input: numbers, text, and etc.
- Model Structure: **Decision Tree, Adaboost**

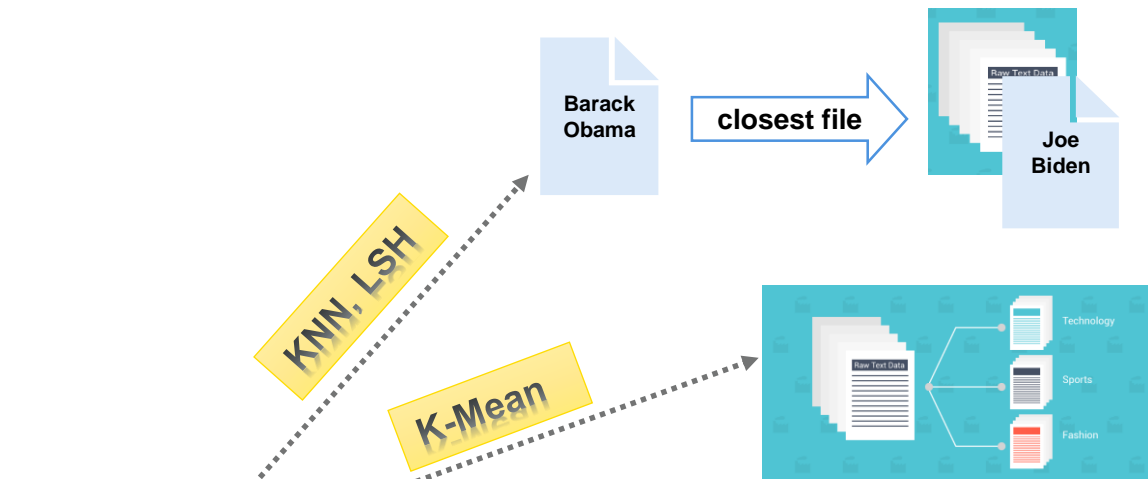
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
id	member_id	loan_id	funded	funded	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_income	is_inc	issue_date	loan_status	payment_url	desc	purpose	title	zip
1E+06	1E+06	5000	5000	4975	36 mo	10.65	162.9	B	B2		10+ yrs	RENT	24000	Verified	201112	Fully Paid	https://	Borrow credit	Computer	8	
1E+06	1E+06	2500	2500	2500	60 mo	15.27	59.83	C	C4	Ryder	< 1 year	RENT	30000	Source	201112	Charged	https://	Borrow car	bike	3	
1E+06	1E+06	2400	2400	2400	36 mo	15.96	84.33	C	C5		10+ yrs	RENT	12252	Not Verified	201112	Fully Paid	https://www.l	small real es	6		
1E+06	1E+06	10000	10000	10000	36 mo	13.49	339.3	C	C1	AIR RES	10+ yrs	RENT	49200	Source	201112	Fully Paid	https://	Borrow other	person	9	
1E+06	1E+06	5000	5000	5000	36 mo	7.9	156.5	A	A4	Veolia	3 years	RENT	36000	Source	201112	Fully Paid	https://www.l	wedding	My wedding	8	
1E+06	1E+06	3000	3000	3000	36 mo	18.64	109.4	E	E1	MKC Acc	9 years	RENT	48000	Source	201112	Fully Paid	https://	Borrow car	Car Door	9	
1E+06	1E+06	5600	5600	5600	60 mo	21.28	152.4	F	F2		4 years	OWN	40000	Source	201112	Charged	https://	Borrow small	Expansion	9	
1E+06	1E+06	5375	5375	5350	60 mo	12.69	121.5	B	B5	Starbucks	< 1 year	RENT	15000	Verified	201112	Charged	https://	Borrow other	Building	7	
1E+06	1E+06	6500	6500	6500	60 mo	14.65	153.5	C	C3	Southwest	5 years	OWN	72000	Not Verified	201112	Fully Paid	https://	Borrow debt_c	High in	8	
1E+06	1E+06	12000	12000	12000	36 mo	12.69	402.5	B	B5	UCLA	10+ yrs	OWN	75000	Source	201112	Fully Paid	https://www.l	debt_c	Consolidation	9	

AdaBoost



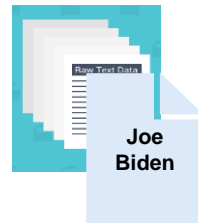
Decision Tree





Barack
Obama

closest file

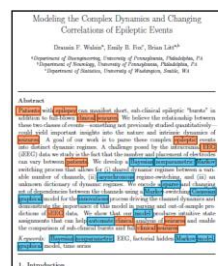


KNN, LSH

K-Mean



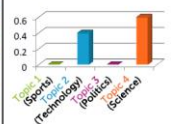
LDA



Mixture Gaussian

LDA is a mixed membership model

Want to discover a set of topics



Arnold Schwarzenegger

From Wikipedia, the free encyclopedia

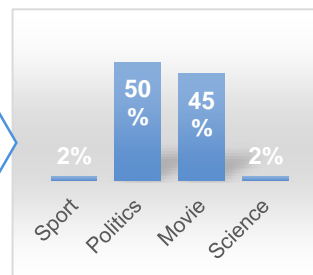
"Schwarzenegger" redirects here. For other people with "The Austrian Oak" redirects here. For the tree, see Quercus.

Arnold Alois Schwarzenegger (/ˈɑːrnɒld ʃwɑːrzənˈɛɡər/)^[?] German professional bodybuilder.^[2] He served as the 38th Governor

Schwarzenegger began lifting weights at the age of 15. He bodybuilding and writing many books and articles on the subject. He is widely considered to be one of the greatest

Schwarzenegger's most notable film role was as the character

Soft assignment



Cluster Wikipedia Documents

The task is to cluster similar documents together and find the topics.

- Project: Cluster Wikipedia Documents
- Data input: text
- Model Structure: Tokenize, bag of word, TF-IDF, KNN, Kmean, LSH, LDA, Guassian distribution
- Problem solved: nearest neighbor, probabilities, cluster

Recommendation System

The algorithm is to build up a movie recommendation system.

- Project: Movie recommendation system
- Model Structure: Collaborative filtering, Content based recommendation

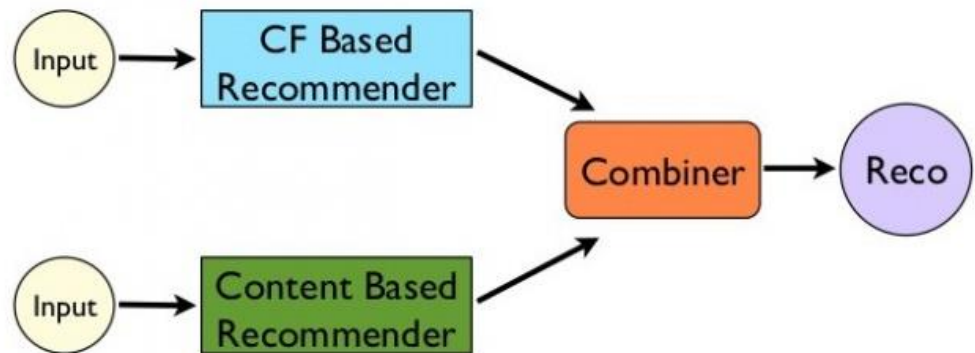
Content based

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)	x_1 (romance)	x_2 (action)
Love at last	5	5	0	0	0.9	0
Romance forever	5	?	?	0	1.0	0.01
Cute puppies of love	?	4	0	?	0.99	
Nonstop car chases	0	0	5	4	0.1	
Swords vs. karate	0	0	5	?	0	

Collaborative filtering

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)
Love at last	5	5	0	0
Romance forever	5	?	?	0
Cute puppies of love	?	4	0	?
Nonstop car chases	0	0	5	4
Swords vs. karate	0	0	5	?

Input



Model

Top recommendations for you:

- Predicting rating 9.0 for movie Titanic (1997)
- Predicting rating 8.9 for movie Star Wars (1977)
- Predicting rating 8.8 for movie Shawshank Redemption, The (1994)
- Predicting rating 8.5 for movie As Good As It Gets (1997)
- Predicting rating 8.5 for movie Good Will Hunting (1997)
- Predicting rating 8.5 for movie Usual Suspects, The (1995)
- Predicting rating 8.5 for movie Schindler's List (1993)



Output

Anomaly Detection

Anomaly Detection

The project is to detect the abnormal data from all data.

- Project: Anomaly detection
- Statistic model: **Univariate Gaussian , Multivariate Gaussian**

Statistic Model

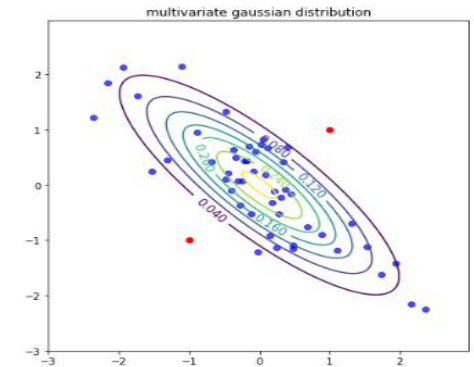
All data

Normal observations



Anomaly

x



Anomaly detection algorithm

1. Choose n features x_i that you think might be indicative of anomalous examples.
2. Fit parameters $\mu_1, \dots, \mu_n, \sigma_1^2, \dots, \sigma_n^2$

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_j^{(i)} \quad \sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_j^{(i)} - \mu_j)^2$$

3. Given new example x , compute $p(x)$:

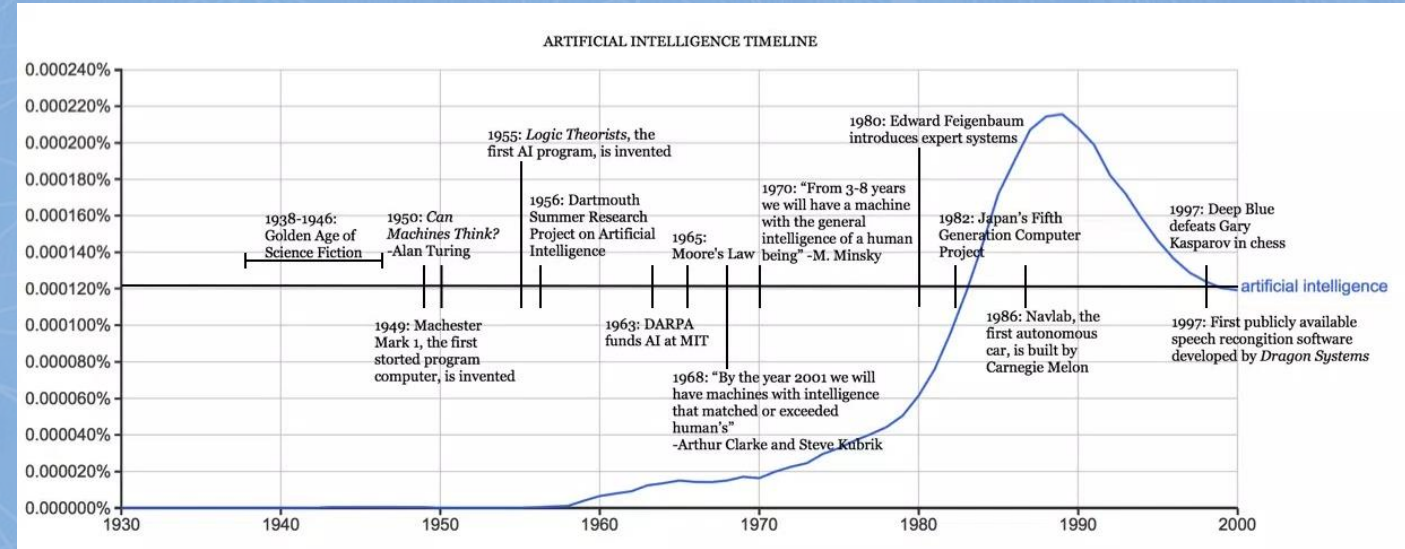
$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right)$$

Anomaly if $p(x) < \varepsilon$

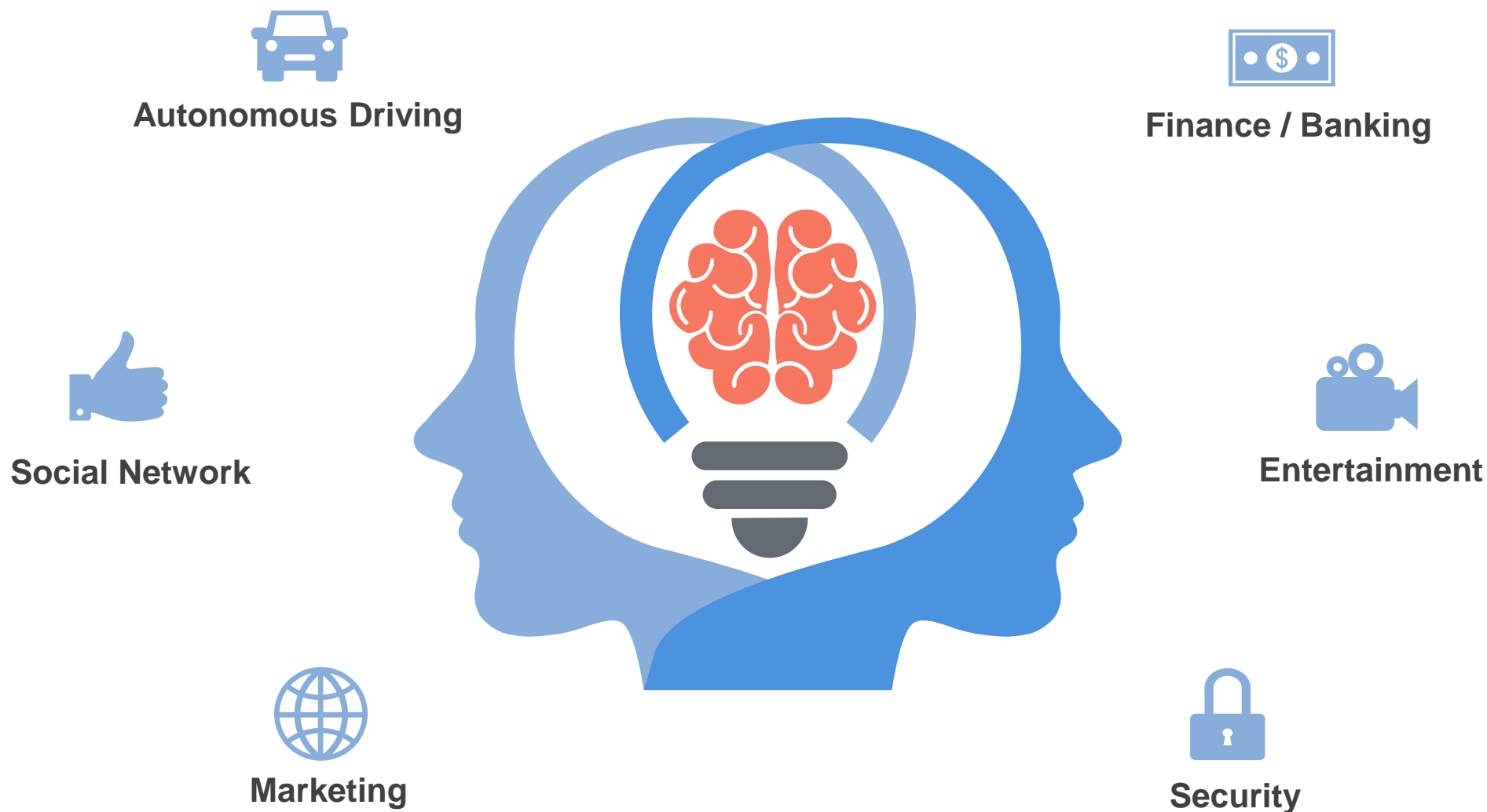
ARTIFICIAL INTELLIGENCE

Artificial Intelligence is Everywhere

We now live in the age of “big data,” an age in which we have the capacity to collect huge sums of information too cumbersome for a person to process. The application of artificial intelligence in this regard has already been quite fruitful in several industries such as technology, banking, marketing, and entertainment. Deploying and exploring in the field will make a big difference in the world in the next twenty years, I cannot wait to be part of it.



Portfolio Summary



The fields are involved in my projects.



Thank You

Feel free to contact me if you
have any questions

yanjialie@gmail.com