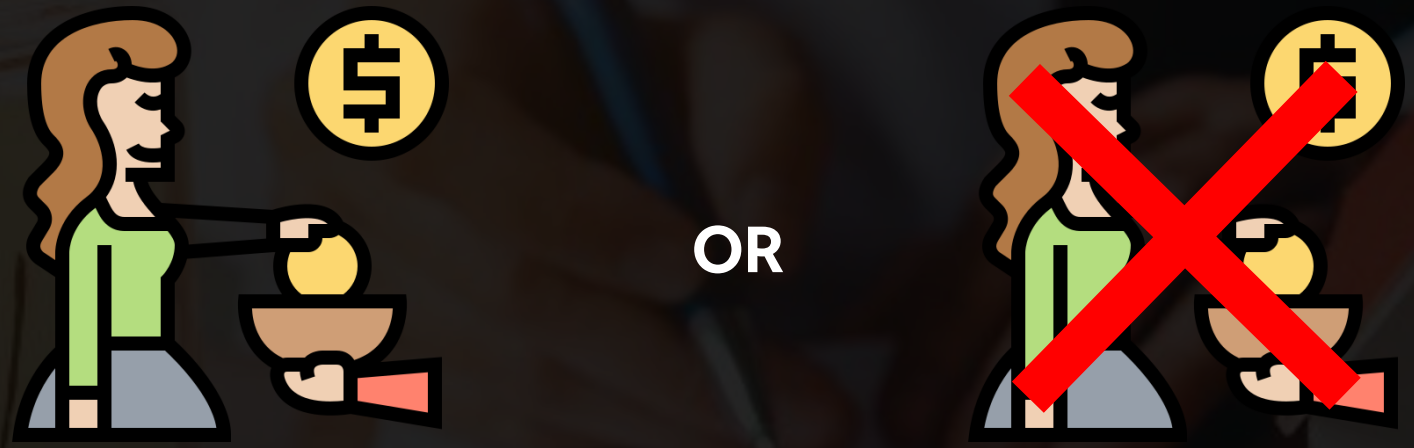
Term Project
Proposal

201631513 Hwang ByungHoon
201635825 Oh SeongWon
201635841 Lee ChaeHyeon
201835437 Kim JinKyung

- INDEX

– Project Title

In Classification

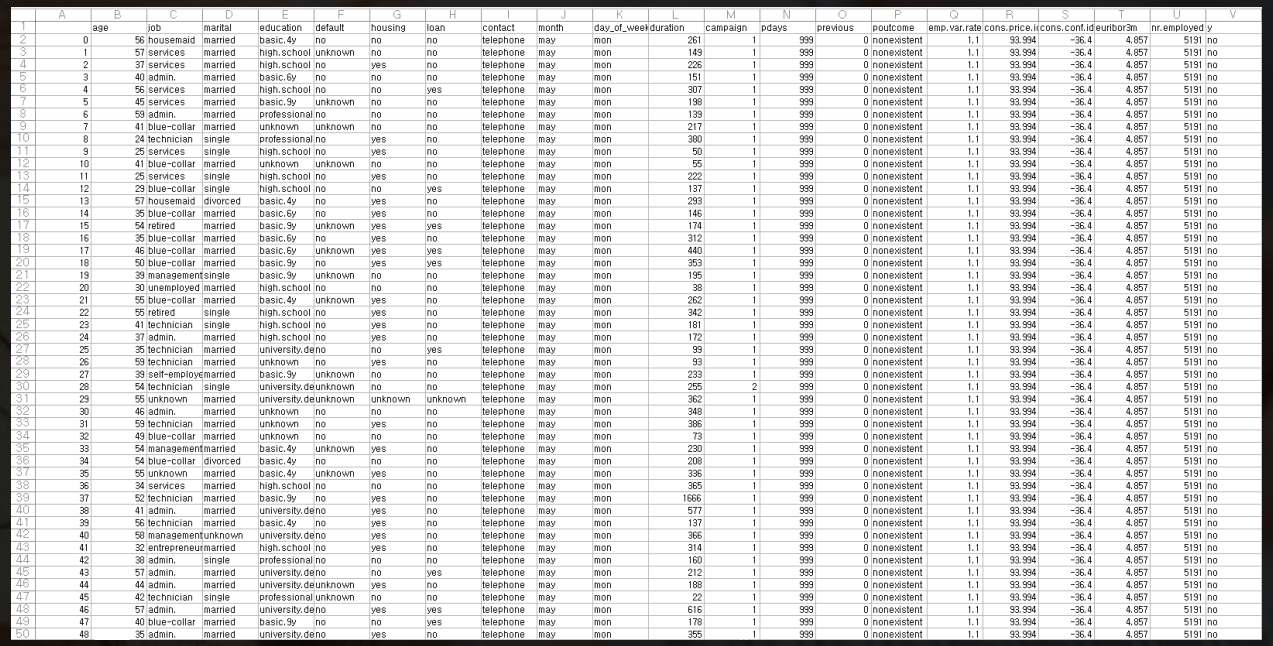


OR

The classification goal is to predict if the client will subscribe a term deposit

- Dataset

In Classification



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays	previous	poutcome	emp.var.rate	cons.price.i	cons.conf.id	euribor3m	nr.employed	y
2	0	56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
3	1	57	services	married	high.school	unknown	no	no	telephone	may	mon	149	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
4	2	37	services	married	high.school	no	yes	no	telephone	may	mon	226	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
5	3	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
6	4	56	services	married	high.school	no	no	yes	telephone	may	mon	307	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
7	5	45	services	married	basic.9y	unknown	no	no	telephone	may	mon	198	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
8	6	59	admin.	married	professional	no	no	no	telephone	may	mon	139	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
9	7	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
10	8	24	technician	single	professional	no	yes	no	telephone	may	mon	380	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
11	9	25	services	single	high.school	no	yes	no	telephone	may	mon	50	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
12	10	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	55	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
13	11	25	services	single	high.school	no	yes	no	telephone	may	mon	222	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
14	12	29	blue-collar	single	high.school	no	no	yes	telephone	may	mon	137	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
15	13	57	housemaid	divorced	basic.4y	no	yes	no	telephone	may	mon	293	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
16	14	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	146	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
17	15	54	retired	married	basic.9y	unknown	yes	yes	telephone	may	mon	174	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
18	16	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	312	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
19	17	46	blue-collar	married	basic.6y	unknown	yes	yes	telephone	may	mon	440	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
20	18	50	blue-collar	married	basic.9y	no	yes	yes	telephone	may	mon	353	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
21	19	39	management	single	basic.9y	unknown	no	no	telephone	may	mon	195	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
22	20	30	unemployed	married	high.school	no	no	no	telephone	may	mon	38	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
23	21	55	blue-collar	married	basic.4y	unknown	yes	no	telephone	may	mon	262	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
24	22	55	retired	single	high.school	no	yes	no	telephone	may	mon	342	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
25	23	41	technician	single	high.school	no	yes	no	telephone	may	mon	181	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
26	24	37	admin.	married	high.school	no	yes	no	telephone	may	mon	172	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
27	25	35	technician	married	university.d	no	no	yes	telephone	may	mon	99	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
28	26	59	technician	married	unknown	no	yes	no	telephone	may	mon	93	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
29	27	39	self-employe	married	basic.9y	unknown	no	no	telephone	may	mon	233	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
30	28	54	technician	single	university.d	unknown	no	no	telephone	may	mon	255	2	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
31	29	55	unknown	married	university.d	unknown	unknown	unknown	telephone	may	mon	362	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
32	30	46	admin.	married	unknown	no	no	no	telephone	may	mon	348	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
33	31	59	technician	married	unknown	no	yes	no	telephone	may	mon	386	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
34	32	49	blue-collar	married	unknown	no	no	no	telephone	may	mon	73	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
35	33	54	management	married	basic.4y	unknown	yes	no	telephone	may	mon	230	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
36	34	54	blue-collar	divorced	basic.4y	no	no	no	telephone	may	mon	208	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
37	35	55	unknown	married	basic.4y	unknown	yes	no	telephone	may	mon	336	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
38	36	34	services	married	high.school	no	no	no	telephone	may	mon	365	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
39	37	52	technician	married	basic.9y	no	yes	no	telephone	may	mon	1666	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
40	38	41	admin.	married	university.d	no	yes	no	telephone	may	mon	577	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
41	39	56	technician	married	basic.4y	no	yes	no	telephone	may	mon	137	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
42	40	58	management	unknown	university.d	no	yes	no	telephone	may	mon	366	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
43	41	32	entrepreneur	married	high.school	no	yes	no	telephone	may	mon	314	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
44	42	38	admin.	single	professional	no	no	no	telephone	may	mon	160	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
45	43	57	admin.	married	university.d	no	no	yes	telephone	may	mon	212	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
46	44	44	admin.	married	university.d	unknown	yes	no	telephone	may	mon	188	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
47	45	42	technician	single	professional	unknown	no	no	telephone	may	mon	22	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
48	46	57	admin.	married	university.d	no	yes	yes	telephone	may	mon	616	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
49	47	40	blue-collar	married	basic.9y	no	no	yes	telephone	may	mon	178	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no
50	48	35	admin.	married	university.d	no	yes	no	telephone	may	mon	355	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no

File
bank-additional-full.csv

Size
21 Columns,
41147 Rows

Dataset Link
https://www.kaggle.com/henriqueyamahata/bank-marketing?select=bank-additional-full.csv

- Dataset

In Classification

Numeric

- Age
- Duration
- Campaign
- Pdays
- Previous
- emp.var.rate
- cons.price.idx
- cons.conf.idx
- euribor3m
- nr.employed

Categorical

- Job
- Marital
- Education
- Default
- Housing
- Loan
- Contact
- month
- day_of_week:
- Poutcome
- Y (target)

"There is no null value in the data"

- Dataset

In Classification

Age (numeric)
→ We think the perception of deposits will vary depending on age.

Job : type of job (categorical)
→ It will have an impact because profits depend on jobs.

Default : has credit in default? (categorical)
→ Those who are credit in default will not hold deposits.

Housing : has housing loan? (categorical)
→ The loan of the house will affect deposit registration.

Loan : has personal loan? (categorical)
→ Individuals' loan status will affect deposit registration.

- Dataset

In Classification

Emp.var.rate : employment variation rate - quarterly indicator (numeric)

Cons.price.idx : consumer price index - monthly indicator (numeric)

Cons.conf.idx : consumer confidence index - monthly indicator (numeric)

Euribor3m : euribor 3 month rate - daily indicator (numeric)

Nr.employed : number of employees - quarterly indicator (numeric)

Economic conditions can affect people's enrollment in deposits.

– Data Preprocessing

In Classification

1. Outlier Data detection

2. Feature Scaling

→ min-max normalization

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

3. Splitting the Training and Test set

→ Cross validation

- Brief Description

In Classification

Algorithm

- KNN
- SVM
- Decision Tree
- Random Forest
- GradientBoostingClassifier
- XGBClassifier
- GaussianNB

Term Project
Proposal

- Project Title

In Clustering

The clustering goal is
to show how each country's population
changes over time.

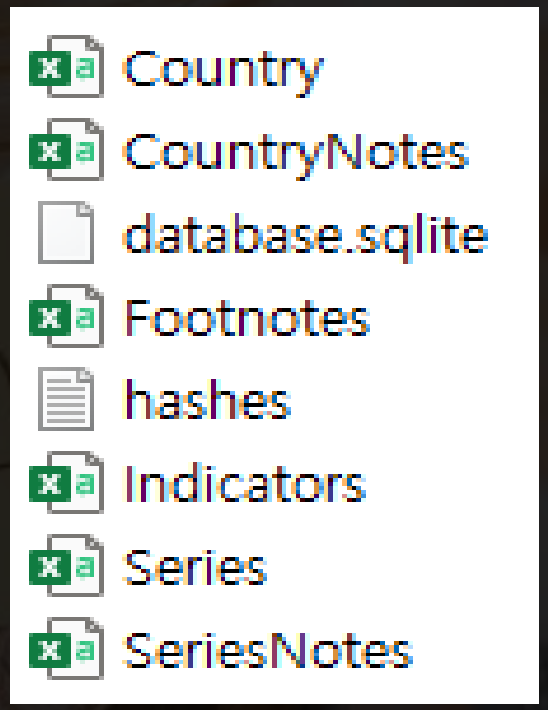
– Project Idea

In Clustering

- By clustering the indicator dataset, we analyze how the population has changed by year.
- It also visualizes these results by displaying them on a map.

- Dataset

In Clustering



Has 6 csv Files

Explore country development indicators from around the world

Dataset Link

https://www.kaggle.com/worldbank/world-development-indicators

– Brief Description

In Clustering

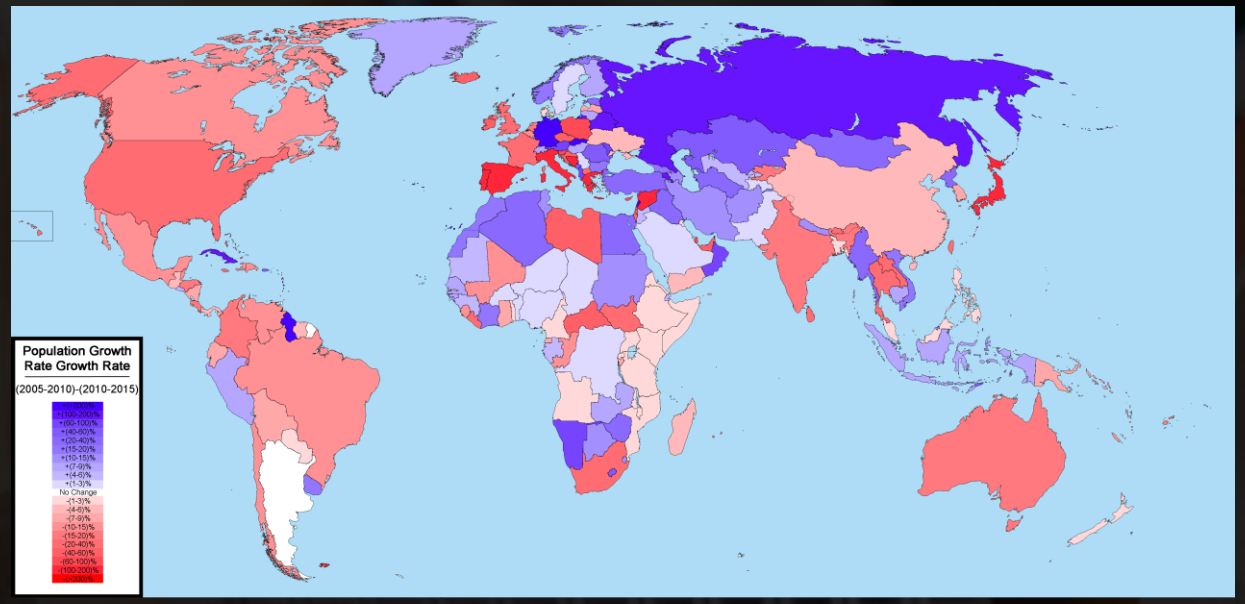
- Use 3 Machine Learning Algorithms – K Means, DBSCAN, EM
  - We will use the best of the three algorithms

Visualization using PCA

Evaluation

– Brief Description

In Clustering



The world's population growth is expressed in colors on the map by year

- Team member

Hwang ByungHoon

Student ID
201631513

Email
byunghoonhwang97
@gmail.com

Oh SeongWon

Student ID
201635825

Email
tjddnjs970325@gmail.com

Lee ChaeHyeon

Student ID
201635841

Email
lch50484084@gmail.com

Kim JinKyung

Student ID:
201835437

Email
lxsz987@gmail.com

- Schedule

Sun	Mon	Tue	Wed	Tur	Fri	Sat
11/1	11/2	11/3	11/4	11/5	11/6	11/7
Data Curation / Data Inspection				Data Preprocessing		
11/8	11/9	11/10	11/11	11/12	11/13	11/14
		Data Analysis (Classification)				
11/15	11/16	11/17	11/18	11/19	11/20	11/21
Data Analysis (Clustering)						
11/22	11/23	11/24	11/25	11/26	11/27	11/29
Evaluation / Deployment					Prepare Final Presentation	
11/29	11/30	12/1	12/2	12/3	12/4	12/5
Prepare Final Presentation			Final Presentation			

Thank You