DMBI MINI Project on Suicide or Suicide Attempts in INDIA in 2001
By
HARSH SHAH

Problem definition, identifying which data mining task is needed.

Problem Statement:-

Suicide In India is a most Critical as well as Crucial thing to be noted,because unlike in other countries Every Hour and Every Day People commit suicide in India due to many reasons like marriage issues,Dowry system,students do due to family pressure or study pressure,Farmers do because they are not even earning a single profit out of what they sell,Businessman Dies due to loss of money in stock market etc and many more reasons.In the year 2001,all these above things were happened together including some additional problems like people commited suicide or attempted suicide due to Job Crises and Market Inflation. About 8Lakh people commit suicide worldwide out of which 20% is from India Only.thus for analysing we use NAÏVE BAYSE Classification Algorithm for Persons commiting/Attempting Suicide in 2001.

Identify and use a standard data mining dataset available for the problem.Some Links for data mining dataset are:- WEKA site,UCI site,KDD site and KDD cup site etc.

Data Exploration:-

The dataset " Suicide or Suicide Attemps in India 2001" was taken from the UCI machine learning repository which comes from Luis Candanedo,luismiguel.candanedoibarra „@‟umons.ac.be,UMONS.

Implement the data mining algorithm of choice.

*Algorithm used for Mining is: NAÏVE BAYES CLASSIFICATION ALGORITHM*

Naïve Bayes is a classification technique based on Bayes theorem with an assumption of independence among predictors.It actually assumes that presence of a particular feature in a class is unrelated to the presence of other feature.For example,a fruit may be considered to be an apple if it is red in color,round in shape and about 3 inches in diameter.This classification Algorithm is easy to build and particularly useful for large dataset along with simplicity.Naive bayes is a highly sophisticated classification method.

Interpret and visualize the results

Targeted data attribute is TOTAL which includes 0 and 1 values which means people commited suicide and attempted suicide respectively in 2001.

ExampleSet (149 examples, 1 special attribute, 6 regular attributes)   Filter (149 / 149 examples): all

Row No.	Total	State	Year	Type_code	Type	Gender	Age_group
44	0	A & N Islands	2001	Causes	Fall in Social ...	Male	0-14
45	0	A & N Islands	2001	Causes	Dowry Dispute	Male	0-14
46	0	A & N Islands	2001	Causes	Ideological C...	Male	0-14
47	0	A & N Islands	2001	Causes	Illegitimate Pr...	Male	0-14
48	1	A & N Islands	2001	Causes	Failure in Exa...	Male	0-14
49	0	A & N Islands	2001	Causes	Bankruptcy or...	Male	0-14
50	0	A & N Islands	2001	Causes	Insanity/Ment...	Male	0-14
51	0	A & N Islands	2001	Causes	Divorce	Male	0-14
52	0	A & N Islands	2001	Causes	Suspected/Illi...	Male	0-14
53	0	A & N Islands	2001	Causes	Suspected/Illi...	Female	15-29
54	0	A & N Islands	2001	Causes	Illegitimate Pr...	Female	15-29
55	1	A & N Islands	2001	Causes	Other Prolon...	Female	15-29
56	1	A & N Islands	2001	Causes	Causes Not k...	Female	15-29
57	0	A & N Islands	2001	Causes	Bankruptcy.or...	Female	15-29
58	0	A & N Islands	2001	Causes	Cancer	Female	15-29

Design Modelling for our Project
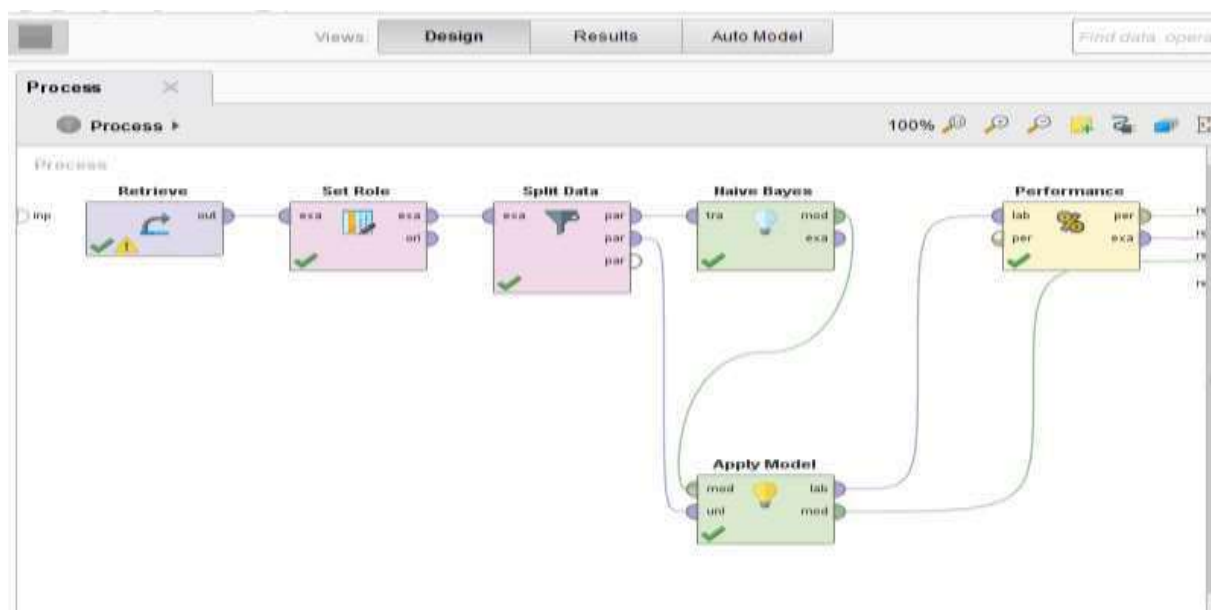
Parameter wise Distribution of an attribute

Attribute	Parameter	1	0
State	value=A & N Islands	1.000	1.000
State	value=unknown	0.000	0.000
Year	mean	2001	2001
Year	standard deviation	0.001	0.001
Type_code	value=Causes	1.000	1.000
Type_code	value=unknown	0.000	0.000
Type	value=Illness (Aids/STD)	0.043	0.039
Type	value=Bankruptcy or Sudden chang...	0.000	0.039
Type	value=Cancellation/Non-Settlement ...	0.000	0.066
Type	value=Physical Abuse (Rape/Incest ...	0.000	0.053
Type	value=Dowry Dispute	0.000	0.013
Type	value=Family Problems	0.129	0.013
Type	value=Ideological Causes/Hero Wo...	0.000	0.039
Type	value=Other Prolonged Illness	0.172	0.000

SimpleDistribution

Distribution model for label attribute Total

Class 1 (0.232)
6 distributions

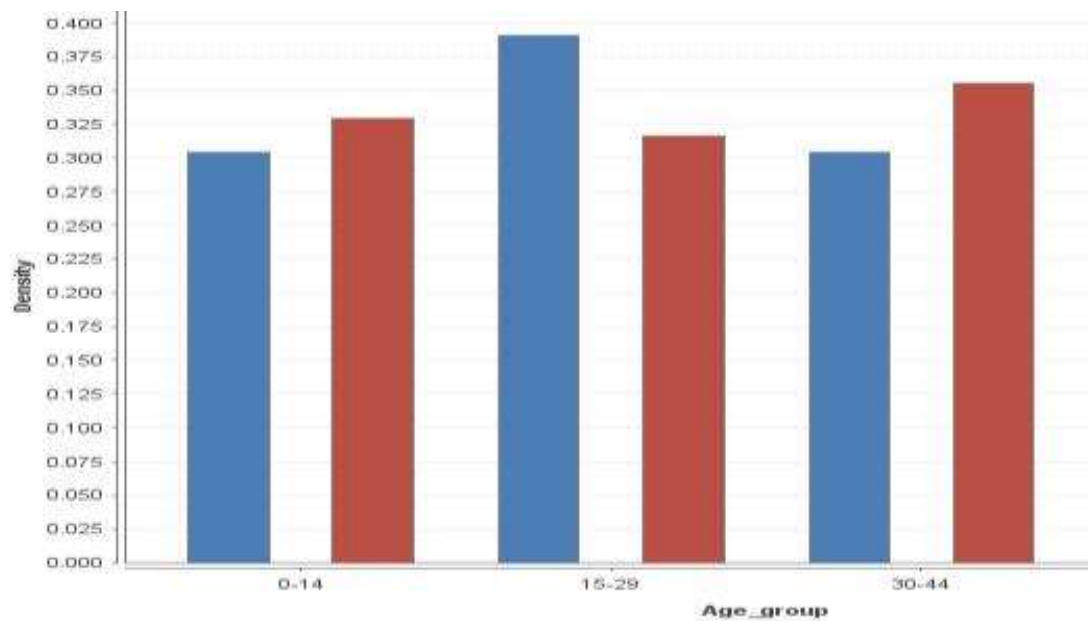Class 0 (0.768)
6 distributions

Classification and Accuracy

classification_error: 20.00%

	true 1	true 0	class precision
pred. 1	3	7	30.00%
pred. 0	3	37	92.50%
class recall	50.00%	84.09%	

accuracy: 80.00%

	true 1	true 0	class precision
pred. 1	3	7	30.00%
pred. 0	3	37	92.50%
class recall	50.00%	84.09%	

Chart of different Age Group and Gender commiting Suicide

Provide clearly the BI decision that is to be taken as a result of mining.

There is about 1 or more suicide happening across India every 1 hour so together we must standup,speakup and advocate for better policies and implementation of resource for mental health.We must view suicide as a condition needing treatment nor as a punishment.

So A multi pronged approach must be taken to decrease the worlds suicide rate.Then we must have Mental Health Education and resources for dealing with symptoms of mental health should be taught and implemented from an early age and they should be motivated.