# Mining High Economic Status of Judokas through Clustering, **KNN and Decision Trees**

## Reena Hooda<sup>1</sup>

<sup>1</sup>(Dept. of Computer Sc. & Applications, Indira Gandhi University, Meerpur, Rewari, Haryana, India)

ABSTRACT: Data mining turned into astimulating, operative, and widely acceptable methodology for analyzing central repository from different perspective even on the same data and attaining knowledge while elaborating various insights through the applicability of obtainable mining tools. The advantage of using these refined and efficient tools are that besides the required information, they aid in highlighting the same data sets in different modes to provide additionalinvestigation. The present paper underlining the implication of one of the finest tools of data mining for examining the data of 164 Judokas and to demonstrate various relationships between different entities, their grouping and pictographic representation of the classification through the kmeans clustering, k-nearest neighbor method and decision tree classification.

Keywords: Clustering, Mining, Decision tree.

#### I. **INTRODUCTION**

Data mining aims at providing the knowledge about the data in terms of their classification, summary, pictorial representation of data sets, hidden facts, prediction about some unknown values, pattern recognition, data distribution etc. The present paper attempts to enhance the received results of 164 Judokas [6], flow of data. interrelationships, and hierarchical distribution of data and partitions. These 164 students belong to a set of 300 students who gathered at Punjab University for Judo Championship. The work had been highlighting the relationship of achievement motivation and social status factors with the position winning& participations of 360 Judokas [6]. The study had been conducted through considering the six factors such as social change, liberalism, social distance, nationalism, social revolution and untouchability to find the overall significance of these factors over the position winning, attitude and motivation[6]. The current paper signifying a more expressive examination by taking this database as an example data set while considering High Economic status containing 164 data sets in it and will be treated as inputs to various operations like modeling and classification, the tools used for this analysis is the RapidMiner 5.3.013[2].

#### **EVALUATION AND IMPLEMENTATION** II.

The very first step of the analysis is to create a local repository by importing the Excel sheet and transforms the attributes types, setting Id and Label field so as to retrieve the example repository into required format. Fig. 1 shows the database retrieval by importing Excel sheets and Fig 2 shows the retrieved example dataset. The repository is stored under the head Local Repository. Next step is to use the decision tree operator to generate a pictorial classification of the example dataset. This has been done through selecting this operator from the left plane under the head Modeling, subhead Classification & Regression [4] then drag & drop it on the process area. The operator shows a problem here as it is mandatory to assign a retrieve object to it that will be assigned to it later. Retrieve operator [2][4]contains the example dataset. Filtration is required to handle the missing values as well as duplicate values in example database to get the final tree classification. The decision tree is a descriptive and unsupervised methodology and even can work for non-numerical values have the biggest advantage to select it for detailed classification and ease of use. Fig.3 shows the decision tree classification.

Han	This witard galdes you to import your data. Step 2: An Excelle can contain multiple sheets. Please select the one you want to import into RapidAliner. Furthermore, yo can mark a range of cells to be loaded. Jedam Low									. <i>7</i> %	This wizard Step 4: Rap RapidMiner also defined	guides you to i dhiner uses s assigns roles I here, Finally,	mport your data. Intengy typed attributes. In this step, you can define the data types of your attributes. Furthermore, to the attributes, defining what they can be used for by the individual operators. These roles can b you can rename attributes or deselect them entrely.						
A	0	c	D	£	F	Ġ	н	i.	j	Belo	ad data	🖉 Quess valu	e types 🖉	Preview use	s only first 100	rows. Date	format		
Sr. No.	Name of Pla	Name of unit. Sex			ons Achiv	Social Citat	Soca Dist	a Liberalis#	1 It Nationalist	7	G	0	6	5			0	G	6
	Amazana Ya	Purpositive M				94 91			20.34	2	<u></u>	(C)	<u></u>	(E)	(C)	(L)	2	2	E .
	Amondation C	Publich 1 tele M					64.37	43.67	577	Br. No.	Name of Pla	Name of uni	Sex	SES HSES	Sports Achiv	Social Chan	Social Dista	Liberalism I	Nationalism
	Tanwaider K	Punjah Linus M				40.16	120 77	17.75	1515	integer .	text *	ted +	binomi. *	integer *	integer *	numeric *	polyno_ *	numeric *	real +
	Navehetan 5	Guru Manak M					97.72		16.32	attribute .	(anity of a	attribute -	real 🖾	(attribute) =	attribute? +	attribute) +	attribute +	atritute -	(attribute) +
	And Kumar	Guru Nanak M						32.14		and the second	Danager Kin	Dunish (Inte	Set	140	20	64.040	05.64	60.000	27.245
	Wrender Sin	Keculahetra M								-	Impress Vi	Punjab Univ	binomin	110	30	6 5 5 0	10.00	45.635	3.010
	Adsul Fam	Dr. EA. Mara M		3 2			7454	25.52	30.15		Amandeen	Punjah Univ	polynom	128	34	67.470	64.27	42 970	67 700
	Sanderp										Jasseinder H	Punjab Univ	data tir	127	36	40 160	120.77	37 710	19 150
	Ren.P.R.	Calcul Unive M				72.49					Navehetan S	Guru Nanak	date	106	34	58.610	97.72	23,760	16.320
	Navjot Singh	Purgab Univ- M					139.12				Anil Kumar	Guru Nanak	time 🔛	143	30	71.500	96.91	32.640	14.510
	Manpreel Sir	Punjab Univ M				58.29		38 1			Wrender Sin	Kurukshetra	ш	152	30	82.810	93.74	67,880	88.510
	Jaswinder S	Punjab Univ M		7. 3							Adsul Ram	Dr. BA. Mara	M	113	24	33.110	74.54	25.520	30.190
	Marcreet	Goru Nanak, M						50.56		6	Sandeep	Guru Nanak	м	119	32	34.470	67.9	15.300	18.950
	Parmod Kun	C.C.S.Univer M								10	RenPR	Calcut Unive	м	161	22	72.490	100.88	35.010	85,700
	Katkar Suth	Dr BA Mara M	10			41.47			35.58	11	Navjot Singh	Punjač Univ	M	111	28	108.100	139.12	35.720	130.320
	Gandeep fir	Korutshetra M								12	Manpreet Si	Punjab Univ	М	104	34	58,290	115.58	38.600	18.670
	Ammderins	Korutshetra M								13	Jaswinder S	Punjab Univ	М	127	36	40.160	120.77	37.740	19,150
	Acil	Guru filanak M					96.91	32.54	1451	2	Harrist	Providences	н.,		40	10.000	70.07	20.055	
20	Priya	Punjab Unie F		2 34			122.82	30.74		2.0 errors							Filmo		Show only er
	Stephnie E. I	Delh Univer F	10			38.04	66.88	40.52		-	ne Caluma		50				CC) Denos		outer part To
	Sneba Tadyi	Punjab Univ F			A STATE	55.05			59.48	- "	ow, column		EU	и.		unginal varue		Mes	2984
	Rekta Runi	Maharishi D. F	10		36		61.5	17.52	60.05										
	Komal Sahi	Punjab Univ F	12	1. 3			122.59	39,79	41.51										
W		CHARLEN IS	m			and a	dor-th_	10.26	199.5						1/2-				1222

Fig.1. Shows the Database retrieval by importing Excel sheets and setting attributes types.

×					//toci	# Repositor	ty/results - I	capioMiner	53.013 @	Arshuman				- P .
Elle Edt	Excess I	Joola 7/em Relo												
16		BAAR			() ()									
2 8	a Duestiene	The Oraciana The		Exampleter	(Retrieve knot)		Completed /R	amous Paulie	ated) 11 1	Achana	Add / Thestan	na (70	-	C III Passantes II M & III N
(C) Duraute	- 0 -	and the Original Co			Presente Prove							-9 (+))		R A A C C H
S) Data vie	* O tota D	HER KINW () FICT KINW () A	avanced Cn	uns () Annos	nona								<b>7</b> -	8
ExampleSe	t (155 examp	les, 2 special attributes, 11 re	gular attribu	utes)					, y	tew Filter (165	/165) all		٠	B C Samples (normal)
Row No.	Sr. No.	Name of un., Name of PL.	Sex	SES HSES	Sports Achi	Social Cha	Social Dista	Liberalism I.	Nationalia	Social Revo	Untouchable	Position/Pa		Configuration and
	1	Punjab Univ. Parveen Kur	м	118	30	64.910	85.540	50.930	27.340	23,210	114,040	191	6	Golf-Testaet mana-ett
1	2	Punjab Univ Amanwer Ya	M	136	34	8.550	19	45.620	3.010	4.840	15.740	151		- If this prove a st)
1	3	Punjao Univ Amandeep 5	M	128	34	67.470	64,270	42.970	67.700	92.970	48.110	161		- III Labor-Negotiations (mma - 21)
6	4	Punjab Univ Jaswinder K	M	127	36	40.160	120.770	37.710	19.150	34.180	95.430	161		- I Mariel Cata store - el)
5	5	Guru Nanak Navehetan S	м	106	34	58.610	97.720	23.760	16.320	25.740	85.220	fut		- U Polynomial (nume - =1)
5	6	Guru Nanak Anil Kumar	M	143	30	71.500	95.910	32.640	14,510	34.060	85.520	tat		- U Riplay-Dat (name - et)
t	7	Kurukshetra Vivender Sin	м	152	30	82.810	93,740	67.880	88.510	44,520	84,320	210		- Is Sonar (nink - v?)
E.	8	Dr. BA. Marar Adaul Ram	M	113	24	33.110	74.540	25.520	30.190	13.080	24.850	-2nd		- Ig transactions (mma - un)
2	9	Guru Nanak Sandeep	M	119	32	34.470	67.900	15.300	18.960	25.310	74 900	2nd		I Contrastes and
10	10	Calcut Unive Ren.P.R.	м	161	22	72.490	100.880	35.010	85,700	66.090	93.190	2nd		I CO
11	11	Punjab Univ: Nacot Singh	М	111	28	108.100	139:120	35.720	130.320	107.710	133.050	218		🖻 🛷 Local Reportion
12	12	Punjab Univ Manpreet Siz	M	104	34	58.290	115.500	38.600	18.670	27.380	115.760	2nd		(8) 🔄 processes (anami)
0	13	Punjab Univ Jaswinder S	M	127	36	40.160	120.770	37.740	19.150	34,180	95.430	2nd		Scole analysis provided, 19415-4
н.	14	Guru Nanak Manpreet	M	111	18	40.980	79.060	60.890	61.690	17.670	57.080	3rd		- 1 KM (and 1, Sec. 5, 53 PM - 18 a
15	15	C.C.S.Univer Parmod Kun	м	101	30	36.820	64.330	33.240	31,190	21.770	78.930	3rd		- U chaster (and 1 - 1 - 3 + 15 + 29 (14 - 1
6	15	Dr. BA Mara Katkar Sudh	M	105	34	41.470	03.440	22.380	35.580	43.960	94,030	3/d		- 10 results (analy - v1, 3-415 TST PSF-1
7	17	Kurukshetral Sandeep Sir	M	113	32	71.790	75.440	54,200	51.730	21,210	59 730	3/d		
18	18	Kurukshetra Amrinder Ins	M	105	24	59.480	77.443	92.920	45.160	68.220	27.050	3/5		
19	19	Guru Nanak Anil	М	543	30	71.500	95.910	32.640	14.510	34.060	85.520	3rd		
20	20	Punjab Univ Priya	F	122	34	82.560	122.820	30.740	20.300	45.660	92.990	tat		
21	21	Dethi Univer Stephnie E 1	F:	101	36.	38.040	65.880	40.520	53.350	81.910	38.160	1st		
12	22	Punjab Univ Sneha Tadyi	F	127	36	55.050	118.140	95.330	59.480	90,210	74.910	fist		
23	23	Maharishi D. Rekha Rani	F	105	28	14.120	61.500	87 520	69.080	65.720	54.650	2nd		
24	24	Punjao Univ Komal Sani	F.	121	36	70.920	122.590	39,790	41.510	76.770	122.760	2nd		
25	25	Punjab Univ Poonam	F	114	32	88.690	115.710	48.790	20.300	45.500	95,770	268		
NA .	- 14	Makasiski P. Busines	8	456	18	30 300	74 105	83.845	45 KAD	61.010	#2.085	Sec	0	<u>e</u> .

Fig.2. Shows the Retrieved Example dataset.

National Conference on Advances in Engineering, Technology & Management (AETM'15)"



Fig. 3. Shows Decision Tree Classification.

To show the classification in a tabular form under the control of the user, clustering operator has been used that applies the k-means method for partitioning. The limitation of this operator is that it works only on non-nominal values that mean if the input database has the polynomial values etc. other than numerical, this operator shows an error. One option is to input the database excluding polynomials or convert polynomials to real or integer, another option is to use a select operator to filter only the selected attributes that has numeric types. This is the fundamental difference between decision tress classifications and clustering that decision tree considers all kind of attributes including nominal values. However, the advantage of clustering over decision tree is that it is supervised methodology whereas decision tree is unsupervised technique. The values of k in kmeans and kNN is user defined i.e. set at 5 in k-means ranges from 0 to 4. The maximum runs value is set at 10 that is a default value and also be set by the user. This way, out of 164 Judokas, 5 clusters have been created containing 45, 38, 10, 23 and 48 items respectively. But if an option Good Start Value has also been considered, then the items will be 37, 10, 16, 48 and 53 in different clusters. User can also tick Add Cluster Attribute as additional attribute in result set to show the different entities belongs to which particular cluster. This attribute can also be added as a Label[1][3][5] in place of cluster in the result dataset. Further advantage of this clustering k-means operator is that the output can be further targeted to a separate file other than the results set through the use of Write Cluster operator. This Write operator can only work with the cluster model output so it is not suitable for kNN as kNN is a predication operator. The output of kNN is directly linked to the result node. Fig.4, Fig.5 and Fig.6shows the different efficaciesk-means operator offers, such as graph view and text views, centroid table, centroid plot view and even clustered result set. It seems that kNN prediction operator doesn't fit for the classification as compared to k-means and decision trees as this operator is used for predicting a target and takes training set as well as example set to get the output and do not show the clear picture of classification.



Fig.4. Shows Text View and Centroid View.

*		Territoria Destrutor	u du			//Loca	A Repository	//esults* - Rapid	Miner 53.013 @ Anshuman		- 0 ×		
Elle Ed	Process	Icols Yew	Help										
		19 m	10	▶		<b>Q</b> (1)							
	Example	Set (Retrieve kr	a) X	1	Example	eSet (Remove	Duplicates)	N	Examplediet (It.ocal Repository/cluster) 🐰		Repositories X 2 4 0		
	Result Overview X				iii bar	roleSet (Retr	rove Duplicate	s (2)). 🖂	P Tree (Decision Tree)		84.9944		
Data V	ew 🔾 Meta	Data View O P	Not View 🔾 A	wanced Cha	rts () Annot	atora			1	6.0	N C Samples must		
ExampleS	et (154 exam	ples, 2 special a	deloutes, 6 re	pular attribute	42				View Fither (164 / 164) at	*	* (§ ce		
Row No.	id	cluster	SES HSES	South Arth	Social Cha	Nationalis	Social Revo	Universitabil	A SALES AND A SALES		8 4 Local Repository		
1	1	riuster 4	118	30	64.013	27.340	23,210	114 040		0	B processes (anani)		
2	2	cluster 0	136	34	6.550	3.010	4.840	15.740			CUSH and of Sectors in		
3	3	cluster_1	128	34	67.470	67.700	92 970	48.110			- If score analysis provide the back and		
4	4	cluster_4	127	36	40.160	19.150	34.180	95.430			- Steening and a second second second		
5	5	duster_4	105	34	58.610	16.320	25.740	86.220					
6	6	cluster_4	143	30	71.500	14.510	34.060	85.520					
7	7	chuster_2	152	30	82.810	88.510	44.520	84.320					
8	8	cluster_0	113	24	33.110	30,190	13.080	24,850					
9	9	cluster_4	119	32	34,470	18.950	25.310	74,900					
10	10	duster_2	161	22	72,490	85.700	65.990	93.190					
11	99	cluster_2	111	28	108.100	130.320	107.710	133.850					
12	12	cluster_4	104	34	58,290	18.670	27.380	115.760					
13	13	cluster_1	111	18	40.980	61,690	17.870	57.080					
14	14	cluster_4	101	30	36.820	31.190	21.770	78.930					
15	15	cluster_4	105	34	41.470	35.580	43.960	94 030					
16	15	duster_1	113	32	71.790	51,730	21,210	59 730					
17	17	cluster_1	105	24	59.400	45.160	68.220	27.050					
10	58	duster_4	122	34	82.560	20,300	45.560	92.990					
19	19	cluster_1	101	36	38.040	53.350	81.910	38.160					
20	20	cluster_1	127	36	55.050	59.490	90.210	74.910					
21	21	chuster_1	108	28	14.120	69.080	65.720	54.650					
22	22	cluster_4	121	36	70.920	41.510	76.770	122.760					
23	23	cluster_4	114	32	88.690	20.300	45.500	95.770					
24	24	duster_1	108	28	39,300	42.540	51.010	62.960			0		
24	26	chooler_A	432	74	73.050	27.030	67.670	124.065		0	<u>U</u>		

Fig.5. Shows Result file containing an additional Cluster attribute.



Fig.6. Shows the Centroid Plot View of the Clusters.

## III. CONCLUSIONS AND FURTHER SCOPE OF WORK

The present paper demonstrated the significance of mining techniques for representing an acute and meticulous investigation of data items in tabular as well as pictorial forms through inputting the primary data of 164 Judokas to show detailed associations among the six given attributes and effectiveness of achievement motivation and High Economic Status over the position winning and partitioning of Judokas. The further scope of the work may include the enrichment of the mining methodologies and even to work with other operators and tools.

### REFERENCES

- [1] https://rapidminer.com/wp-content/uploads/2013/10/RapidMiner\_RapidMinerInAcademicUse\_en.pdf.
- [2] https://rapidminer.com/
- [3] http://sourceforge.net/projects/rapidminer/
- [4] https://rapidminer.com/wp-content/uploads/2014/10/RapidMiner-v6-user-manual.pdf
- [5] http://www.slideshare.net/dataminingtools/rapidminer-introduction-to-datamining
- [6] D. Dhaka, *Relationship of Motivation and Attitude with Socio Economic Status and Performance of Judokas, Department of Physical Education*, MaharshiDayanand University, Rohtak. 2011.