Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DISCOVER COMPATIBILITY: MACHINE LEARNING WAY

[1] Er. PRITAM R. AHIRE, [2] Dr. PREETI MULAY
SYMBIOSIS INSTITUDE OF TECHNOLOGY, PUNE.
E-mail: [1]pritam.ahire@sitpune.edu.in, [2] preeti.mulay@sitpune.edu.in

ABSTRACT

To live in harmony, it is advisable to have awareness of matching among various individual behaviors. Naïve Bays algorithm with extensions is executed on Zodiac Signs and Name Style datasets to develop a research system. This system is implemented to understand how compatible an individual is based on not only zodiac signs but numerology also. Numerology is a science which suggests different naming style, with unique spelling of person. Every spelling is supported with special set of characteristics. These naming style characteristics are matched with zodiac signs to find similarity index values, which gives compatibility about various relations including family, professional and friend.

Keywords: *Zodiac Signs; Compatibility; Numerology; Similarity Index.*

1. INTRODUCTION

It is observed through various research outcomes [8][18] that compatibility of two individuals mostly depend on their zodiac signs and name style. Numerology plays a very important role along with characteristics of various zodiac signs. If in a relation say friend, boss, family etc. has the matching Sun or Moon sign, then working environment become positive and in turn productivity increases. To overcome the blind belief on astrology, Information Technology based algorithms are used extensively with extensions, to prove this system scientifically. TF-IDF, Naïve Bays, Artificial Neural Network (ANN) along with techniques like stop-word removal, data-parsing, pre-processing, ranking and stemming are used along with.

With this system, a great helping hand is provided to realize the importance of compatibility. Sun, Moon signs and spelling style of proper-noun does affect behavior of an individual with others. While working on this research system, it is observed that same person may carry same or different zodiac signs. These zodiac signs for an individual could be two different extremes for ex. Sun Sign=Leo and Moon Sign=Pisces, one is king of jungle and other completely immersed in water, two extreme mediums indeed, for the same individual. Then the question arises, what is the exact personality of that individual? To answer this question, this research included one additional layer called name-style / name-spelling. Every similar-sounding spelling may have different characteristics for example Priti and Preethi. The focus of this research is to locate similarity index among Sun, Moon and Name-spelling characteristics. Differences are not neglected completely but importance is given to similarity index only, in this research. Figure 1 shows the mind map of system, including ideas and expected output.

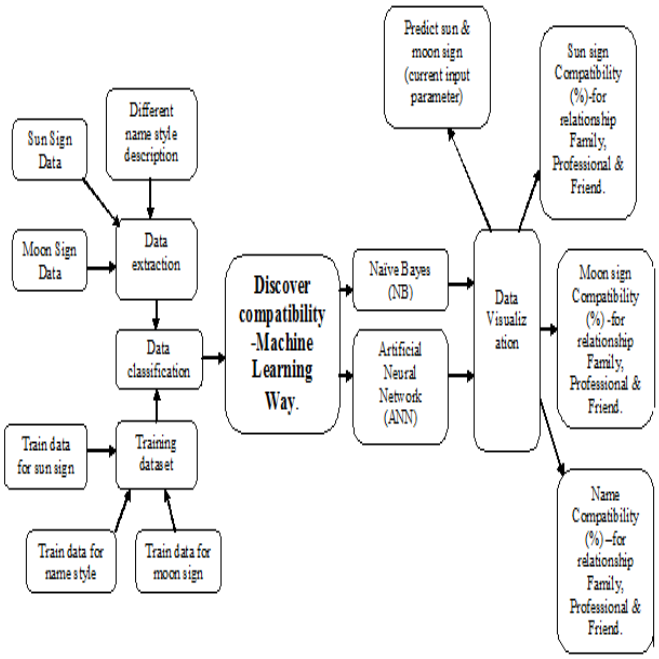

Figure 1: Mind map of the system

As shown in figure 1, this research system is layered in three parts, pre-processing, compatibility check and analyzing + validating results. These

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

three layers can be termed as data-extraction, data-classification and validated visualization [2][9][19]. In data-extraction phase, compiled zodiac and name-style characteristics with similar-sounding / pronounced proper-nouns, is processed according to required relationship status. Being a generalized system, implementation started with taking input from user as date-of-birth, name and gender of person, as it is a well-known fact that characteristics of boys and girls differ.

Data-classification, the supervised paradigm [9] includes finding similarity and differences among Sun, Moon and Name, using Naïve Bays / ANN approach[4] [20][22]. The reason behind choosing these algorithms: they are proved good classifiers among others, giving concrete results, for this research work, having combination of text-mining, prediction and similarity-index related tasks.

Data-visualization and validation helps to confirm the required outcome is correct and complete, hybrid concept based on numerology + astrology.

2. LITERATURE REVIEW

Data Collection and Pre-processing

Data is collected from various validated zodiac sign & numerology sources available on web. System used and manipulate data when required. Pre-processing operations like Data Parsing, Stop word removal operation, Stemming, TF-IDF algorithm, Ranked keyword is carried out in research.

This system parse [15] text description related to zodiac sign & name styles, to compute similarity index & differences among them. After data parsing Stop word removal operation performed. Stemming is important operation in system in the stage of dataset training [14]. Stemming shows root word from available dataset, for because of that user should easy to identify and select word for training & again it make easy to mapping in same word appear in different dataset. System used TF-IDF [5] [17] for the easier identification & selection of word, required for training & mapping from different data. Keyword ranking [7] gives the importance or priority to word in available summarized dataset. System used ranking operation for training dataset, it become easy to identify and select important words which are used for further training operations. In Data Visualization System show the result like similarities index and differences between zodiac sign related to particular relationship between two individual people & is impact with the print spelling of name. Also show the various style of name spell in their impact on zodiac sign. System visualized data in such a way that it shows compatibility between two people with related to particular relationship between them.

3. MODEL ARCHITECHTURE

This system Discover compatibility of relation between two people, their zodiac sign & name description. The matching between them is calculated using two algorithms are NB result given in paper & ANN is future extension. Data is available in different databases for performing specific operation like related to sun sign, moon sign, & name spell type (i.e. naming conventions) & related to synonyms. As demonstrate in fig 2.Model Architecture, the pre-processing includes few activity like data parsing, stop word removal, stemming, cover TF-IDF rule[5], keyword ranking after these steps system achieve summarized dataset, which includes root words & ranked words, using these words system reach to final result. After these step training is applied to available dataset i.e. apply training is the process of select the particular word which is compatible for relationship, and as per these selected words for relationship ,the training is done automatically & efficiently. So in training says, system must select compatible word for particular relationship between two people. After applying training, system have two trained model which find compatibility, & they are using Naive Bayes algorithm [1] [3] [10] [21] i.e. Naive Bayes trained model and using ANN algorithm [11] [12] [13] [16] i.e. ANN trained model.

As show in figure 2, system accept input parameter system accepts input from user is name, date of birth and moon sign, now as per the input parameter system find out its sun and moon sign, and retrieve data from database, so after that retrieve important keyword and the recognized the data and find the compatible between two person for particular relationship for particular name spell.

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

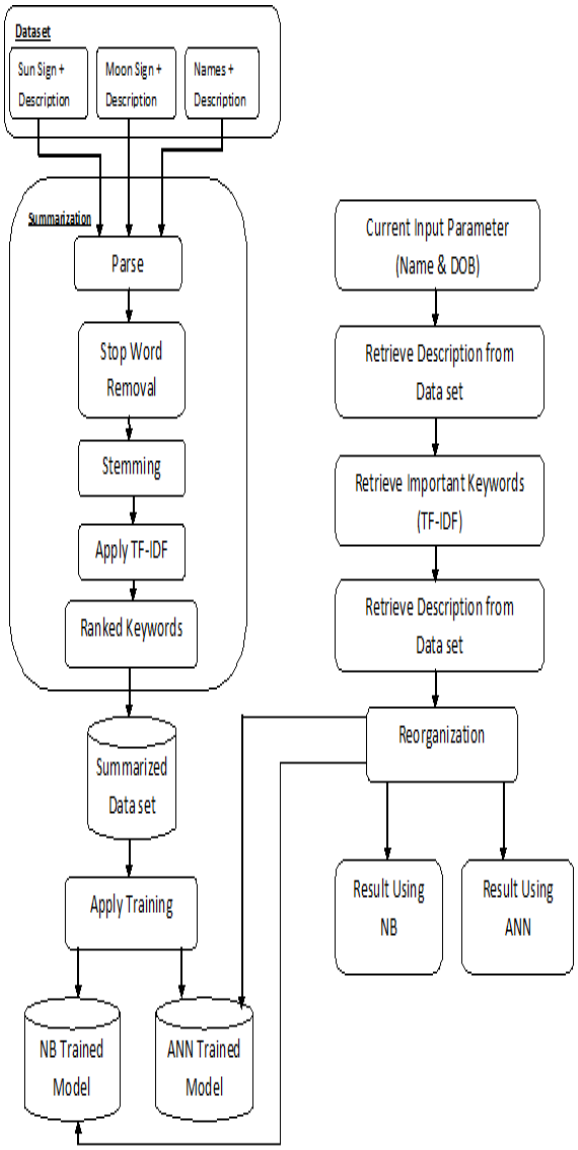

Figure 2: Model Architecture Of Discover Compatibility System.

4. IMPLEMENTATION OF ALGORITHM

Proposed Algorithm to Discover Compatibility:

Discover compatibility: Machine Learning Way:

Compatibility matching algorithm is performing operation using both NB & ANN machine learning algorithm. System shows manage both results coming from NB & ANN for analyzing the result for which algorithm gives frequent and concrete result for input parameter.

Compatibility Matching Algorithm- Machine Learning Methods:

Input db: sun_sign_dataset,
moon_sign_dataset,
name_discription_dataset,
relation_characteristics

Input: name, date of birth.

Output: sun sign, moon sign, sun sign prediction, Moon sign prediction, name prediction, overall result, and percentage of prediction for compatibility.

Step 1: Select Manage relationship words.

Step 2: Add words in particular relationship which user want to train, similarly remove word if user does not word for further training process.
[User should stored some important word in every Relationship, which is frequency occurred in description in signs or name style, so it become easy for training data set]

Step 3: Manage Stop words used for summarization.

Step 4: Manage synonyms [used for efficient train data set].

Step 5: Manage name data set, import available name and sign data set, and select **Generate Dataset.**
[When user selects **Generate data set** system apply **stemmer, stop word removal, TF-IDF algorithm & ranking keyword** to the available description of various parameter i.e. zodiac signs & name].
[System automatically trains the data set using selected words which are important in particular relationship (Step-2), likewise all dataset for sun sign, moon sign, and name description is trained.]
[Train dataset have various important words and there compatibility for particular relationship]

Step 6: Now, select **compatibility checker**.

Step 7: Give input as name & date of birth, and select the relationship for which user want to check compatibility.

Step 8: As per input, system fetch the data from train dataset, for sun

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	sign, moon sign & name styles description.
Step 9:	System shows sun sign, moon sign, sun sign prediction, moon sign prediction, name prediction and overall prediction. System also shows the percentage of matching and mismatching.
Step 10:	Store the similarity index in database for further mapping.
Step 11:	Mapping find out for the compatibility result for using **Naïve Bayes (NB) Algorithm.**
Step 12:	Similar operation are perform for validate and analyze the result using **Artificial Neural Network (ANN) Algorithm.**
Step 13:	Comparing both algorithm result & then analyze the frequency of algorithm, for validate output for using accurate train model.

5. RESULT:

5.1 Results for Compatibility

NO.	Name	Compatibility	Relationship: Father- Child / Husband- Wife / Brother & Sister — Family	Relationship: Boss-Emp / Colleague / Partners — Professional	Relationship — Friend
1	Anand	Sun Sign	50.00%	99.01%	98%
		Moon Sign	50.00%	64.55%	54%
	Goraksh	Name Style	42.27%	37%	14.14%
2	Anand	Sun Sign	52.92%	100%	57%
		Moon Sign	52.92%	75%	28%
	Pritam	Name Style	15.81%	11%	8.00%
3	Anand	Sun Sign	50.00%	99.01%	98%
		Moon Sign	50.00%	64.55%	54%
	Sagar	Name Style	42.27%	37%	14.14%
4	Preety	Sun Sign	52.92%	98.55%	98.25%
		Moon Sign	52.92%	63.96%	60.30%
	Priti	Name Style	51%	68%	53%
5	Preety	Sun Sign	42.64%	99.55%	95.35%
		Moon Sign	42.64%	70.71%	47.67%
	Preeti	Name Style	51%	71%	43%
6	Pritam	Sun Sign	50%	99.01%	98.02%
		Moon Sign	50%	64.55%	54.01%
	Preetam	Name Style	17.32%	30.00%	30.00%
7	Pritam	Sun Sign	48.99%	99.12%	97.99%
		Moon Sign	48.99%	66.33%	52.92%
	Pritham	Name Style	17.32%	30.00%	30.00%
8	Pritam	Sun Sign	48.99%	99.12%	97.99%
		Moon Sign	48.99%	66.33%	52.92%
	Preeti	Name Style	26.46%	37.42%	14.14%
9	Om	Sun Sign	42.64%	99.55%	95.35%
		Moon Sign	42.64%	70.71%	47.67%
	Shivaji	Name Style	0.00%	0.00%	0.00%
10	Komal	Sun Sign	52.92%	100.00%	56.57%
		Moon Sign	52.92%	74.83%	28.28%
	Sneha	Name Style	0.00%	0.00%	0.00%

Figure Fig 3: Result Table For Compatibility Matching Between Two Persons.

Results show the compatibility between two persons for certain relationship like Family (e.g. relation such as father-child (son or daughter), husband-wife, brother- sisters, cousins etc.), Professional (e.g. relations such as Boss-Employee, colleagues, business partners etc.) and relation

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

friend, using characteristics of sun & moon sign, and description of name in various styles. System accepts input from user like name and DOB for two persons, using name system extracts moon sign of persons & using DOB system extract sun sign of person. As per accepted current input parameter system gets sun & moon sign of person, the system extract data for particular sign form dataset, after that mapping the two characteristic and then final result is computed.

As show in table user takes various styles of input for analysis lets discuss briefly.

1. Result No. 1, 2 & 3 : System analyze the result keeping one name constant and changing second, like user keep "anand" name as it is, and change the second input as "Goraksh", "Pritam", "Sagar" for taking various output for analysis.
2. Result No. 4 & 5: System analyzes result for different spell style for similar pronounce. Keep "Preety" same for both result and changing second input spell as "Priti"& "Preeti". System shows output both result does not show similar results i.e. less compatibility between them by changing name spell (show in dark red color mark).
3. Result No. 6 & 7: System analyzes result for different spell style for similar pronounce. Keep "Pritam" same for both result and changing second input spell as "Preetam" & "Pritham". Both show similar results i.e. maximum compatibility between them by changing name spell, for name style (show in Dark green color mark), and system show nearby results for sun & moon compatibility (show in pink color mark).
4. Result No.8: For analysis purpose here system takes results for similar signs (attracted sign) i.e. both persons have similar sun & moon sign (show in green faint green color mark).(**Results for similar signs).**
5. Result No.9 & 10: For analysis purpose here system takes results for exactly opposite signs i.e. repelled signs. Result 9 shows repelled signs between two persons name. Result 10 shows repelled of sign between sun and moon signs. (**Result for two extreme sign/ repelled signs).**

- **5.2 Train dataset for Sun Signs:**

Managing Relationship Word For "Family"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Helpful	1	1	0	1	1	1	1	1	1	0	1	1
2	Understandable	1	0	0	0	1	1	1	1	1	1	0	0
3	Loving	1	0	0	0	1	0	1	1	1	1	0	0
4	Heart	0	0	1	1	1	0	1	1	1	1	0	0
5	Loneliness	0	1	1	1	0	0	1	1	0	1	1	1
6	Wilful	1	1	1	1	0	1	1	1	0	1	1	1
7	Especially	1	1	1	0	0	1	1	1	0	0	1	1
8	Cool headed	1	1	1	1	1	1	1	1	0	1	1	1
9	Emotional	1	1	1	1	1	1	1	1	1	1	1	1
10	careful	1	1	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 4: Training Dataset Of Sun Sign For The Family Relationship.

Managing Relationship Word For "Professional"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Directs	1	1	0	1	1	1	1	1	1	1	1	1
2	Supervise	1	0	0	1	1	1	1	1	1	1	0	0
3	Motivated	1	0	0	1	1	0	1	1	1	1	0	0
4	Impulsive	0	0	1	0	1	0	1	1	1	1	1	0
5	Self-reliant	0	1	1	0	0	0	1	1	0	1	1	1
6	Spontaneous	1	1	1	0	0	1	1	1	0	0	1	1
7	Excitable	1	1	1	0	0	1	1	1	0	0	1	1
8	Vocation	1	1	1	1	0	1	1	1	0	0	0	1
9	Occult	1	1	1	0	0	1	1	1	1	0	0	1
10	Mysticism	1	1	1	1	1	1	1	1	1	1	1	1
11	Self-disciplined	0	1	1	1	1	1	1	1	1	1	1	1
12	Subjective	1	1	1	1	1	1	1	1	1	1	1	1
13	Superficially	1	1	1	1	1	1	1	1	1	1	1	1
14	Insecure	1	1	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 5: Training Dataset Of Sun Sign For Professional Relationship.

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Managing Relationship Word For "Friend"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Kind	1	1	0	1	1	1	1	1	1	1	1	1
2	Helpful	1	0	0	1	1	1	1	1	1	1	1	0
3	Responsible	1	0	0	1	1	0	1	1	1	1	1	0
4	Wonderfully	1	0	1	1	1	0	1	1	0	1	0	1
5	Occasional	1	0	1	0	0	1	1	1	0	0	1	1
6	Subjective	1	0	1	0	0	1	1	1	0	0	0	1
7	Emotions	0	1	1	0	1	1	1	1	1	1	1	1
8	Friendship	1	1	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 6: Training Dataset Of Sun Sign For Friend Relationship.

Above figures represent the train data set of sun sign for all 12 zodiac sign, with its some fix characteristic and its presences and absence in particular zodiac sign for sun sign. Sun sign is manipulate from person DOB, as per user mention DOB system map the exact zodiac sign and the get exact result. Dataset is train automatically as per presence and absence of particular word in available dataset for selected sign for getting appropriate result.

- **5.3 Train dataset for Moon sign:**

Managing Relationship Word For "Family"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Helpful	1	1	1	1	1	1	1	0	1	1	1	1
2	Understandable	1	1	1	1	1	1	1	0	0	0	1	1
3	Loving	1	1	1	1	0	1	1	0	1	1	1	1
4	Heart	1	1	1	1	0	1	1	1	1	1	1	0
5	Loneliness	1	1	0	1	1	1	1	1	1	1	0	1
6	Wilful	1	1	1	1	0	1	1	1	0	1	0	1
7	Especially	1	1	1	1	1	1	1	1	0	0	0	1
8	Cool headed	1	1	1	1	1	1	1	1	0	1	1	1
9	Emotional	1	0	1	1	1	1	1	1	1	1	1	1
10	careful	1	0	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 7: Training Dataset Of Moon Sign For Family Relationship.

Managing Relationship Word For "Professional"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Directs	1	1	1	1	1	1	1	0	1	1	1	1
2	Supervise	1	1	1	1	1	1	1	0	0	0	1	1
3	Motivated	1	1	1	1	0	1	1	1	1	1	1	1
4	Impulsive	1	1	0	1	1	1	1	1	1	1	0	0
5	Self-reliant	1	1	1	1	1	1	1	1	1	0	0	0
6	Spontaneous	1	1	1	1	1	1	1	1	1	1	0	1
7	Excitable	1	1	1	1	1	1	1	1	0	1	0	1
8	Vocation	1	1	0	1	0	1	1	1	0	1	0	1
9	Occult	1	1	1	1	0	1	1	1	1	0	1	1
10	Mysticism	1	1	1	1	0	1	1	1	1	1	1	0
11	Self-disciplined	1	1	1	1	1	1	1	1	1	1	1	1
12	Subjective	1	1	0	1	1	1	1	1	1	0	1	1
13	Superficially	1	0	1	1	1	1	1	1	1	1	1	1
14	Insecure	1	0	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 8: Training Dataset Of Moon Sign For Professional Relationship.

Managing Relationship Word For "Friend"

		Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricon	Aquarius	Pisces
1	Kind	1	1	1	1	1	1	1	0	1	1	1	1
2	Helpful	1	1	1	1	1	1	1	0	1	1	1	1
3	Responsible	1	1	1	1	1	1	1	0	1	1	1	1
4	Wonderfully	1	1	0	1	1	1	1	1	1	1	1	0
5	Occasional	1	1	1	1	1	1	1	1	0	1	1	1
6	Subjective	1	1	1	1	0	1	1	1	0	1	1	1
7	Emotions	1	0	0	1	0	1	1	1	1	1	1	0
8	Friendship	1	0	1	1	1	1	1	1	1	1	1	1

1= Present & 0= Absent

Figure 9: Training Dataset Of Moon Sign For Friend Relationship.

Above figures represent the train data set of moon sign for all 12 zodiac sign, with its some fix characteristic and its presences and absence in particular zodiac sign for moon sign. Moon sign is manipulate from person name, as per name the characteristics is mapped for accurate moon sign extracted by system. Dataset is train automatically as per presence and absence of particular word in available dataset for selected sign for getting appropriate result.

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. RESULT GRAPHS

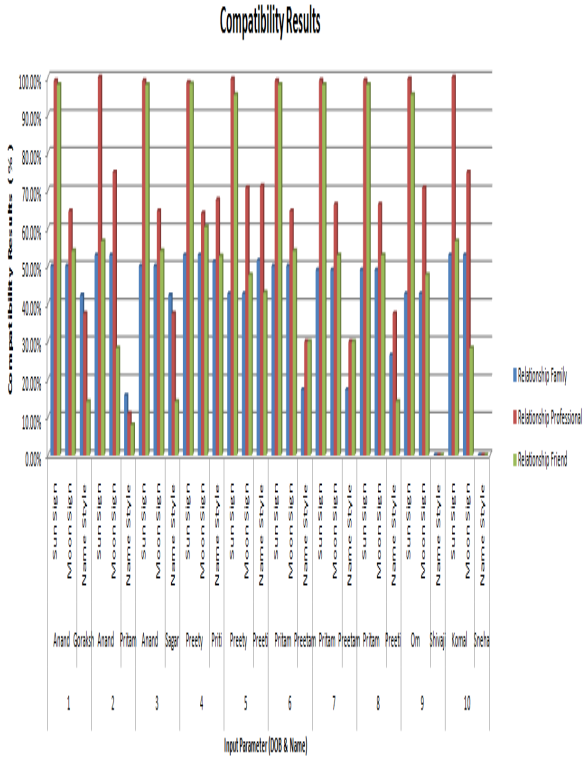

Fig 10: Result Graph

Graph shows the results coming from various styles of input, scale X-axis denotes input parameter i.e. two names and relationships between them, related to Y-axis shows Compatibility result in percentage. Because of graph it becomes easy to analyze results of sun & moon sign parameters. Various colors styles show results given by system. E.g. result 1, "Anand" & "Goraksh", (i) shows result for sun sign "relationship Family = 50%","relationship Professional= 99.01%" &"relationship Friend =98%", similarly (ii) shows results for moon sign "relationship Family = 50%","relationship Professional = 64.55%" & "relationship Friend =54%", and finally (iii) shows result for Name Style "relationship Family = 42.27%","relationship Professional= 37%" &"relationship Friend =14.14%". Likewise all results are plotted in graph, to get better analyze results.

7. VALIDATING THE RESULT

NO.	Name	Compatibility	Relationship		
			Family	Professional	Friend
	Validation Measure -->		Recall	Precision	F-score
1	Anand	Sun Sign	0.416666667	0.5	0.454545455
		Moon Sign	0.416666667	0.5	0.454545455
	Goraksh	Name Style	0.352240845	0.422689014	0.38426274
2	Anand	Sun Sign	0.440958552	0.529150262	0.481045693
		Moon Sign	0.440958552	0.529150262	0.481045693
	Pritam	Name Style	0.131761569	0.158113883	0.143739894
3	Anand	Sun Sign	0.416666667	0.5	0.454545455
		Moon Sign	0.416666667	0.5	0.454545455
	Sagar	Name Style	0.352240845	0.422689014	0.38426274
4	Preety	Sun Sign	0.440958552	0.529150262	0.481045693
		Moon Sign	0.440958552	0.529150262	0.481045693
	Priti	Name Style	0.425930403	0.511116484	0.464651349
5	Preety	Sun Sign	0.355334527	0.426401433	0.387637666
		Moon Sign	0.355334527	0.426401433	0.387637666
	Preeti	Name Style	0.428812609	0.514575131	0.467795574
6	Pritam	Sun Sign	0.416666667	0.5	0.454545455
		Moon Sign	0.416666667	0.5	0.454545455
	Preetam	Name Style	0.144337567	0.173205081	0.157459164
7	Pritam	Sun Sign	0.40824829	0.489897949	0.445361771
		Moon Sign	0.40824829	0.489897949	0.445361771
	Pritham	Name Style	0.144337567	0.173205081	0.157459164
8	Pritam	Sun Sign	0.40824829	0.489897949	0.445361771
		Moon Sign	0.40824829	0.489897949	0.445361771
	Preeti	Name Style	0.220479276	0.264575131	0.240522846
9	Om	Sun Sign	0.355334527	0.426401433	0.387637666
		Moon Sign	0.355334527	0.426401433	0.387637666
	Shivaji	Name Style	0	0	0
10	Komal	Sun Sign	0.440958552	0.529150262	0.481045693
		Moon Sign	0.440958552	0.529150262	0.481045693
	Sneha	Name Style	0	0	0

Fig 11: Validating Results Values (Using Recall, Precision & F-Score.

To validate the system, several experiments are conducted using the system results for certain input parameter. Validation is done by using evaluating values of precision, recall & F-score [6], & then compares these values with system result table. Validation table shows experiments comparison in result 1, 2 & 3 shows compatibility between various styles of inputs. Result 4 & 5 shows less compatibility between two persons using different name style but similar pronounce (show in red color mark). Similarly for result 5 & 6 show high compatibility between two persons using different name style but similar pronounce (show in green color mark). Result 8 shows results for similar signs (attracted signs). And result 9 & 10 shows result for two extreme sign/ repelled signs. Now, comparing result table & validation table for analyzing purposed it is proved that both results i.e. less & high compatibility is show similar output in result

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and validation tables. From these output the results are validates and show similar compatibility output for same input parameter in situated in result table.

8. CONCLUSION AND FUTURE WORK

This research system proved a unique combination of Zodiac signs in-harmony with Numerology using Machine Learning algorithm. Three different relations categories were considered (Family, Profession, and Friend) to prove the results. Productivity of individual definitely increases if compatibility details are known in advance and not learnt by experience. Analyzing the result coming from both the algorithm for finding out which algorithm generates frequently result for input parameter. The incremental learning system built in this research proved to be beyond matchmaking, as life is full of various relations.

The incremental learning system built in this research proved to be beyond matchmaking, as life is full of various relations. This system proved every individual should be aware of people around them, either at home or work. Extension of system to add the significant role relations user should manage relationship for further operation; also user will analyze result by giving gender with name. I.e. system will analyze result for same name but variation in gender, e.g. checking result for "Sameer" as a boy, and also for "Sameer" as a girl.

REFRENCES:

[1] R. Najafi, Mohsen Afsharchi, "Network Intrusion Detection Using Tree Augmented Naive-Bayes", CICIS,IASBS, University of Zanjan,IEEE 2012.

[2] Rudy AG. Gultom, Riri Fitri Sari, Bagio Budiardjo "Implementing Web Data Extraction and Making Mashup with Xtractorz",Dept.of Electrical Engineering, University of Indonesia, Vol 1, pp 4244-4791,2010 IEEE.

[3] Shadab Adam Pattekari, Asma Parveen, "PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES", International journal of advance computer & mathematical science, India,Vol 3, Issue 3,pp 290-294, 2012.

[4] Alexander J. Stimpson and Mary L. Cumming(Senior Member, IEEE), "Assessing Intervention Timing in Computer-Based Education Using Machine Learning Algorithms", Vol 2, IEEE-2014.

[5] Ari Aulia Hakim, Alva Erwin, Kho I Eng, Maulahikmah Galinium, Wahyu Muliady, "Automated Document Classification for News Article in Bahasa Indonesia based on Term Frequency Inverse Document Frequency (TF-IDF) Approach", 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2014.

[6] Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu, "An Ontology-Based Text Mining Method to Cluster Proposals for Research Project Selection", Department of Information Systems, City University of Hong Kong, Renmin University of China,IEEE TRANSACTIONS ON SYSTEMS, MAY 2012.

[7]Anusha Bagalkotkar, Ashesh khandelwal, Shivam Pandey, Sowmya Komaths," A novel technique for efficient text document summarization as a services", third international conference in advance un computing & communications,IEEE-2013.

[8] PA Kulkarni, P Mulay, " Evolve systems using incremental clustering approach", Evolving Systems 4 (2), 71-85.

[9] P Mulay, PA Kulkarni, "Knowledge augmentation via incremental clustering: new technology for effective knowledge management", International Journal of Business Information Systems 12 (1), 68-87.

[10] Levent Koc , Thomas A. Mazzuchi, Shahram Sarkani, "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier", Expert Systems with Applications, pp 13492–13500, 2012.

[11] Rodrigo Moraes, Joao Francisco Valiati, Wilson P. Gaviao Neto, "Document-level sentiment classification: An empirical comparison Between SVM and ANN", Expert Systems with Applications,pp 621-633, 2013.

[12] Othman o-khalifa, Md. Khorshed Alam, Aisha Hassan Abdalla, "An Evaluation on Offline Signature Verification using Artificial Neural Network Approach", International Conference on computing electrical and electronic engineering(ICCEEE), Vol 1, 2013.

[13] Chang Sim Vui, Gan Kim Soon, Chin Kim On, and Rayner Alfred, Patricia Anthony, "A Review of Stock Market Prediction with Artificial Neural Network (ANN)", IEEE International Conference on Control System, Computing and Engineering, Vol 1,pp 477-482, 29 Nov 2013.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14] Wikipedia, "Stemming", http://en.wikipedia.org/wiki/Stemming, last accessed 2 September 2014.
[15] Wikipedia, "Data Parsing", http://en.wikipedia.org/wiki/ Data Parsing, last modified on 19 January 2016.
[16] Anil K. Jain,Jianchang Mao," Artificial Neural Network: A Tutorial", 0018-9162/96,1996 IEEE.
[17] Xing Huang, Qing Wu, "Micro-blog Commercial Word Extraction Based On Improved TF-IDF Algorithm",Vol 1, IEEE 2013.
[18] Abdollah Ansari, Azuraliza Abu Bakar," A Comparative Study of Three Artificial Intelligence Techniques: Genetic Algorithm, Neural Network, and Fuzzy Logic, on Scheduling Problem", 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, Vol 1, pp 31-36, IEEE-2014.
[19] Murat Can Ganiz, Melike Tutkan, Selim Akyokus, "A Novel Classifier Based on Meaning for Text Classification", IEEE 2015.
[20] GuoQiang, "An Effective Algorithm for Improving the Performance of Naive Bayes for Text Classification", Second International Conference on Computer Research and Development, pp 699-701, IEEE 2010.
[21] Liangxiao Jiang, Harry Zhang, and Zhihua Cai, "A Novel Bayes Model: Hidden Naive Bayes", IEEE transactions on knowledge and data engineering, VOL. 21, NO. 10, IEEE 2009.
[22] Lijun Zhou, Zhiyi Yu, Jie Lin, Shikai Zhu, Weijing Shi, Haijie Zhou, Kunpeng Song, and Xiaoyang Zeng," Acceleration of Naive-Bayes Algorithm on Multicore Processor for Massive Text Classification", International Symposium on Integrated Circuits (ISIC), pp 344-347, IEEE 2014.