ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

Language analysis based on speakers of the population using Orange Data Mining

[1] Manikanta V
[1] SRP-Technical in CIIL and Research Scholar in Bharathiar University, Coimbatore, Tamilnadu, India
[1] ciil.manikanta@gmail.com

Abstract: Language is essential through communicate the people for their feelings, thoughts and ideas. Language is one type of system to conventional spoken, set of sounds and written symbols. In our country is diversity of cultural, social, religious, economical, political, traditional and number of variety is more compare to any other country. The number of language divided into scheduled language, non scheduled language and other minor languages. In Indian language are classified into 121 languages from the included constitution eighth schedule 22 languages and non eighth schedule 99 languages. Based on Census of population can be easily visualize the data how language families are distributed, scheduled languages and comparative analysis explored effective and efficient manner using Visual analytic tools.

Keywords: Visual analysis, Orange, Languages, Dravidian, Aryan, Indo-European.

1. INTRODUCTION

Languages are categorized into family based on South Asian region. From this Indian languages are classified into 5 family categories.

1. INDO-EUROPEAN:
The family belongs to regions of Southwest and South Asia with Europe spoken languages.

(a) INDO-ARYAN
From 1800-1500 BCE, ethnolinguistic of south asia group people are diverse speak to the Indo-aryan languages.
There are 21 INDO-ARYAN languages [15]
1. Assamese (S), 2. Bengali(S), 3. Bhili/Bhilodi, 4. Bishnupuriya, 5. Dogri(S) 6. Gujarati(S), 7. Halabi,
8. Hindi(S), 9. Kashmiri(S), 10. Khandeshi, 11. Konkani(S), 12. Lahnda. 13. Maithili(S), 14. Marathi(S), 15. Nepali(S), 16. Odia(S), 17. Punjabi(S), 18. Sanskrit(S), 19. Shina, 20. Sindhi(S), 21. Urdu(S),

(b) IRANIAN
Community of ethno-religious belongs to Zoroastrians migrated in the 19th and 20th centuries from Iran to British [15]
1. Afghani/Kabuli/Pashto

(c) GERMANIC
Northern Germany and southern Scandinavia of Germanic tribes are lived.
1. English.

2. DRAVIDIAN

Dravidian Family languages are commonly used Telugu, Malayalam, Kananda and Tamil. But languages are made up the language of the Dravidian using 70 individual languages. The Dravidian languages are divided into literary language and nonliterary language. The literary languages are Kannada, Tamil, Malayalam and Telugu. Nonliterary languages are completely Proto-South Dravidian languages, South-Central Dravidian languages, Central Dravidian Languages and North Dravidian languages. Proto-South Dravidian is divided into Proto-Tamil-Kannada and Tulu language. Tulu is the Karnataka of Dakshin kannada district and the Kerala of Cannanore district. The Proto-Tamil-Kannada is divided into Proto-Tamil-Toda and Pre-Kannada type of languages. Pre-Kannada type of languages are belongs to Badaga and Kannada language. Proto-Tamil-Toda is divide into Proto-Tamil-Kodagu and Proto-kota-Toda. The Proto-Kota-Toda languages are belongs to kota and Toda. Proto-Tamil-Kodagu is divided into Proto-Tamil-Malayalm and Kodagu language. Kodagu language is belongs to Coorg district. Proto-Tamil-Malayalam is divided into Pre-Tamil and Malayalam language. Pre-Tamil languages are belongs to Tamil and Irula language. South-Central Dravian langauges are divided into Proto-Telugu and Proto-Gondi-Kui. The Proto-Telugu languages are belongs to Savara and Telugu language. Proto-Gondi-Kui languages are divided into Proto-Gondi, Proto-Kui-Kuvi languages, Pengo and Manda language.

Proto-Gondi languages are belongs to Gondi and Konda languages. Proto-Kui-Kuvi languages are belongs to Kui and Kuvi languages. The Proto-Central Dravidian are divided into Proto-Kolami-Naki and Proto-Gadaba-Parji. Proto-Kolami-Naki languages are belongs to Kolami andNaki. Proto-Gadaba-Parji languages are belongs to Gadaba and Parji languages. North Dravidian languages are divided into Proto-Kurukh-Malto and Proto-Brahui. Proto-Kurukh-Malto languages are belongs to Kurukh and Malto Language. Proto-Brahui is belongs to Brahui language.

There are 17 DRAVIDIAN languages [15]

1. Coorgi/Kodagu, 2.Gondi, 3. Jatapu, 4. Kannada(S), 5. Khond/Kondh,, 6. Kisan, 7. Kolami, 8. Konda, 9.Koya, 10. Kui, 11. Kurukh/Oraon, 12. Malayalam(S), 13. Malto, 14. Parji, 15. Tamil(S), 16. Telugu(S), 17.Tulu.

3. AUSTRO-ASIATIC

Austro-Asiatic Languages are divided into Khasi-Khmuic, (Nuclear) Non-Khmer and Munda. Munda languages are belongs to Remo, Savara, Kharia-Juang, Korku.

There are 14 AUSTRO-ASIATIC languages [15]

1. Bhumij, 2. Gadaba, 3. Ho, 4. Juang, 5. Kharia, 6. Khasi, 7. Koda/Kora, 8. Korku, 9. Korwa, 10. Munda, 11.Mundari, 12. Nicobarese, 13. Santali(S) 14.Savara.

4. TIBETO-BURMESE

The Indian languages consists Southeast Asia and other some parts of East Asia and South Asia which include three types of languages Tibetan, Chinese and Burmese. Tibeto-burmese is a Sino-Tibetan language family.

There are 66 TIBETO-BURMESE languages [15]

1. Adi, 2. Anal, 3. Angami, 4. Ao, 5. Balti, 6. Bhotia, 7. Bodo (S), 8. Chakesang, 9. Chakru/Chokri, 10.
Chang, 11. Deori, 12. Dimasa, 13. Gangte, 14. Garo,15. Halam, 16. Hmar, 17. Kabui, 18. Karbi/Mikir, 19. Khezha, 20. Khiemnungan, 21.Kinnauri, 22. Koch, 23. Kom, 24. Konyak, 25. Kuki, 26. Ladakhi, 27. Lahauli, 28.Lakher, 29. Lalung, 30. Lepcha, 31. Liangmei, 32. Limbu, 33.Lotha, 34. Lushai/Mizo, 35. Manipuri(S), 36. Mao, 37. Maram, 38. Maring, 39. Miri/Mishing, 40. Mishmi, 41. Mogh, 42. Monpa, 43. Nissi/Dafla, 44. Nocte, 45. Paite, 46. Pawi, 47. Phom, 48. Pochury, 49. Rabha, 50. Rai, 51.Rengma, 52. Sangtam, 53. Sema, 54. Sherpa, 55. Tamang 56. Tangkhul, 57. Tangsa, 58. Thado, 59. Tibetan, 60. Tripuri, 61. Vaiphei, 62. Wancho, 63. Yimchungre, 64. Zeliang, 65. Zemi, 66. Zou.

5. SEMITO-HAMITIC

The language family consists of around 300 languages belongs to North Africa, West Asia, Sahel, etc.

1. Arabic/Arbi

II. VISUAL ANALYSIS TOOLS

Process of data based cleaning, modeling and transforming are mainly focused on to discover the knowledge and sharing information with meaningful. The information is occupied the conclusions, decision making studies and inference of methods. Some of the free analysis software are available such as R Software Environment, Tableau Public, Microsoft R, Shogun, Orange Data Mining, DataMelt, TANAGRA, Arcadia Data Instant, RapidMiner Starter, Edition, NodeXL, Visual Understanding Enviornment, Lavastorm Analytics Engine, ELKI, Scilab, PAW, Trifacta, Dataiku DSS Community, ROOT, ITALASSI, Fluentd, ITALASSI, DataPrepartor, Scikit-learn, NumPy, SciPy, Google Fusion Tables, DataWrangler, NetworkX, SymPy, Watson Studio, Julia, Datacracker, Data Applied, EasyReg, OpenRefine, Ipython, FreeMat, Massive Online Analysis, Matplotlib, jMatLab, etc are analysis software to explored the knowledge.

ORANGE Data Mining

ORANGE is open source software for Interactive Visualization data and Visual Programming. The data are explored in intelligent way of representation, distributions, scatter plots, box plots, decision trees, clustering, heatmaps and other built in functionalities like Projections and MDS. Regression based classification based on scores, ranking, etc. Many datasets are available to learn visually and Graphic representation with several nodes to compare the data. Clever Graphic explanations are explored while executing any type of data with complete view of information with effectively.

III. VISUAL REPRESENTATIONS AND METHODS

Datasets are mainly imported from the CSV file, TAB file, etc using File option in Orange data mining software. Orange Data Mining provides the tools of Data, visualize, Model, Evaluate, Unsupervised of data tools, etc. Use the File node from Data for importing the TAB file. Datasets are consists of Discrete, Continuous, etc type of variable. From this classified the data based on our prediction value depending upon the tools of the software. Link File node to k-means node and display the plots in visually using ‘visually tool’ from Scatter Plot and MDS (Fig 1). Display the result in Data Table from ‘Data Tool’. Tree analysis of language family is using ‘Tree’ and ‘Tree Viewer’ method to describe the strength and weakness of the language family. Below Tables and Figures are

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

representing the visual results. Scatter Plot provides the X axis and Y axis between two values, color of the groups, shape of any attribute, size, label of fields to view visually, symbol size, opacity, jittering, show color regions, show legend, show gridlines, show all data on mouse hover, show regression line and zoom options. MDS plot is a multidimensional scalability to generate the view of clustered information is visually. MDS consists of PCA, Randomize, Jitter to provide visual information of graphics and tools of color, shape, size, label, symbol size, opacity, jittering, show similar pairs, show color regions and show legend are main functionality of plot. The MDS technique computes a low-dimensional projection of points. The distance matrix used to a plane fitted to given distance between points. Silhouette score is a measured based on how similar pair of data are present in the input.

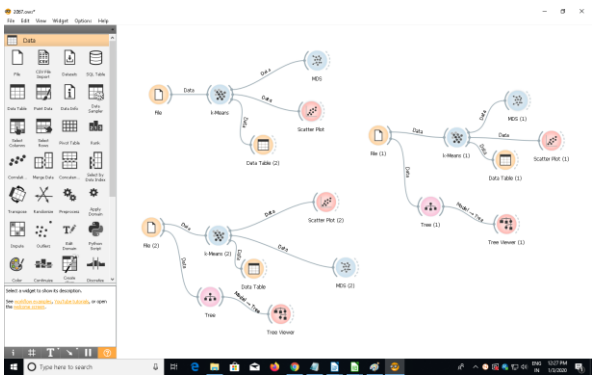


Fig 1. ORANGE DATA MINING TOOLS

The datasets are mapped with languages in various regions of census population for speaker strength.

Language	Count	Family
ASSAMESE	15311351	1
BENGALI	97237669	1
BODO	1482929	4
DOGRI	2596767	1
GUJARATI	55492554	1
HINDI	528347193	1
KANNADA	43706512	2
KASHMIRI	6797587	1
KONKANI	2256502	1
MAITHILI	13583464	1
MALAYALAM	34838819	2
MANIPURI	1761079	4
MARATHI	83026680	1
NEPALI	2926168	1
ODIA	37521324	1
PUNJABI	33124726	1
SANSKRIT	24821	1
SANTALI	7368192	3
SINDHI	2772264	4
TAMIL	69026881	2
TELUGU	81127740	2
URDU	50772631	1

Table 1: Family of Scheduled languages and Population in 2011

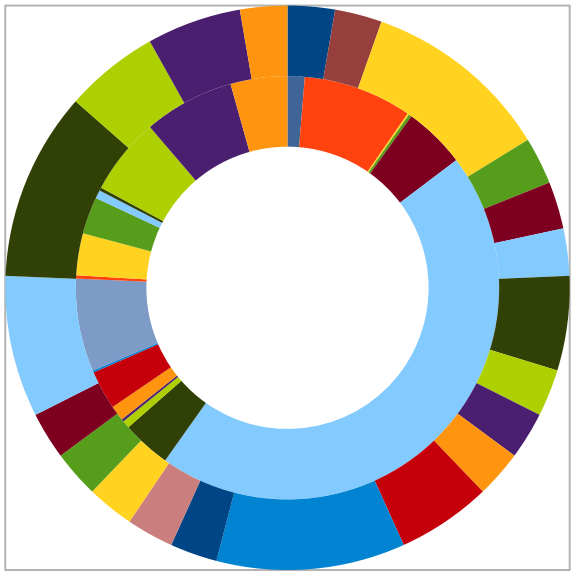


Fig 2. Scheduled languages and Population in 2011

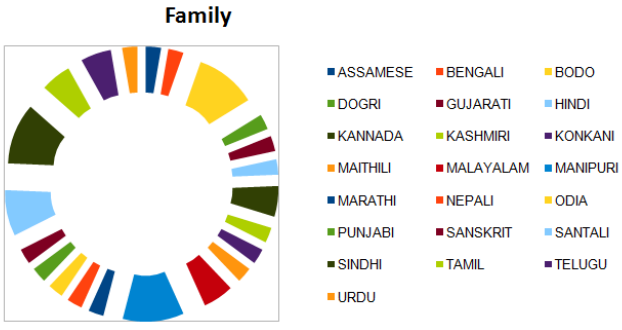


Fig 3. Scheduled languages and Population in 2011 from Family wise

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

Language	Count	Family
AFGHANI/KABULI/PASHTO	21677	1
BHILI/BHILODI	10413637	1
BISHNUPURIYA	79646	1
ENGLISH	259678	1
HALABI	766297	1
KHANDESHI	1860236	1
LAHNDA	108791	1
SHINA	32247	1
COORGI/KODAGU	113857	2
GONDI	2984453	2
JATAPU	20028	2
KHOND/KONDH	155548	2
KISAN	206100	2
KOLAMI	128451	2
KONDA	60699	2
KOYA	407423	2
KUI	941488	2
KURUKH/ORAON	1988350	2
MALTO	234991	2
PARJI	52349	2
TULU	1846427	2
BHUMIJ	27506	3
GADABA	40976	3
HO	1421418	3
JUANG	30378	3
KHARIA	297614	3
KHASI	1431344	3
KODA/KORA	47268	3
KORKU	727133	3
KORWA	28453	3
MUNDA	505922	3
MUNDARI	1128228	3
NICOBARESE	29099	3
SAVARA	409549	3
ADI	248834	4
ANAL	27217	4
ANGAMI	152796	4
AO	260008	4
BALTI	13774	4
BHOTIA	229954	4
CHAKHESANG	19846	4
CHAKRU/CHOKRI	91216	4
CHANG	66852	4
DEORI	32376	4
DIMASA	137184	4
GANGTE	16542	4
GARO	1145323	4
HALAM	38915	4
HMAR	98988	4
KABUI	122931	4
KARBI/MIKIR	528503	4
KHEZHA	41625	4
KHIEMNUNGAN	61983	4
KINNAURI	83561	4
KOCH	36434	4
KOM	15108	4
KONYAK	244477	4
KUKI	83968	4
LADAKHI	14952	4
LAHAULI	11574	4
LAKHER	42429	4
LALUNG	33921	4
LEPCHA	47331	4
LIANGMEI	49811	4
LIMBU	40835	4
LOTHA	179467	4
LUSHAI/MIZO	830846	4
MAO	240205	4
MARAM	32460	4
MARING	25814	4
MIRI/MISHING	629954	4
MISHMI	44100	4
MOGH	36665	4
MONPA	13703	4
NISSI/DAFLA	406532	4
NOCTE	30839	4
PAITE	79507	4
PAWI	28639	4
PHOM	54416	4
POCHURY	21654	4
RABHA	139986	4
RAI	15644	4
RENGMA	65328	4
SANGTAM	76000	4
SEMA	10802	4
SHERPA	16012	4
TAMANG	20154	4
TANGKHUL	187276	4

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

TANGSA	38624	4
THADO	229340	4
TIBETAN	182685	4
TRIPURI	1011294	4
VAIPHEI	42748	4
WANCHO	59154	4
YIMCHUNGRE	83259	4
ZELIANG	63529	4
ZEMI	50925	4
ZOU	26545	4
ARABIC/ARBI	54947	5

Table 2: Non Scheduled Languages of Census of Population from Family wise

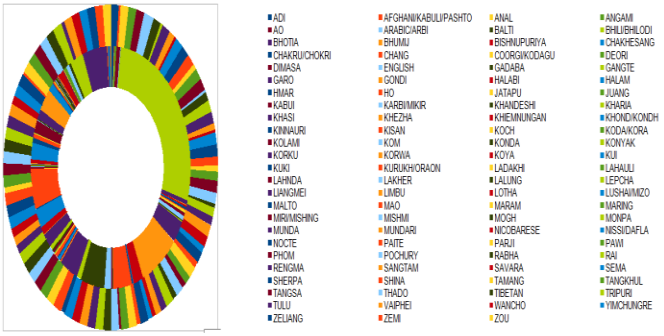


Fig 4: Non-Scheduled languages and Population in 2011

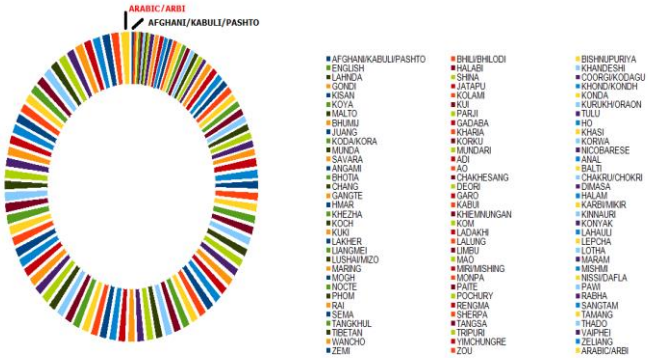


Fig 5: Non-Scheduled languages and Population in 2011 from Family wise

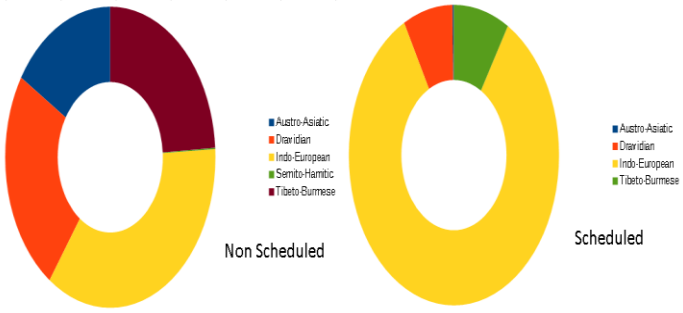


Fig 6: Family wise comparison between Non Scheduled and Scheduled Language in 2011

State	Scheduled	Non Scheduled
1 Jammu & Kashmir	12199484	341818
2 Himachal Pradesh	6721050	143552
3 Punjab	27703349	39989
4 Chandigarh#	1053474	1976
5 Uttarakhand	10029461	56831
6 Haryana	25306294	45168
7 NCT of Delhi#	16757013	30928
8 Rajasthan	64873819	3674618
9 Uttar Pradesh	199770172	42169
10 Bihar	103844271	255181
11 Sikkim	449632	160945
12 Arunachal Pradesh	385707	998020
13 Nagaland	234781	1743721
14 Manipur	1662202	1193592
15 Mizoram	135506	961700
16 Tripura	2563639	1110278
17 Meghalaya	434757	2532132
18 Assam	28952961	2252615
19 West Bengal	90793259	482856
20 Jharkhand	29734312	3253822
21 Odisha	38741904	3232314
22 Chhattisgarh	23162552	2382646
23 Madhya Pradesh	67280520	5346289

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

24 Gujarat	59380062	1059630
25 Daman & Diu#	240716	2531
26 Dadra & Nagar Haveli#	214203	129506
27 Maharashtra	107293455	5080878
28 Andhra Pradesh	83815597	765180
29 Karnataka	58940444	2154853
30 Goa	1441498	17047
31 Lakshadweep#	55,151	9322
32 Kerala	33263039	143022
33 Tamil Nadu	72098315	48715
34 Puducherry#	1246854	1099
35 Andaman & Nicobar Islands#	324400	56181

Table 3: State wise mapping of scheduled and non scheduled census of population

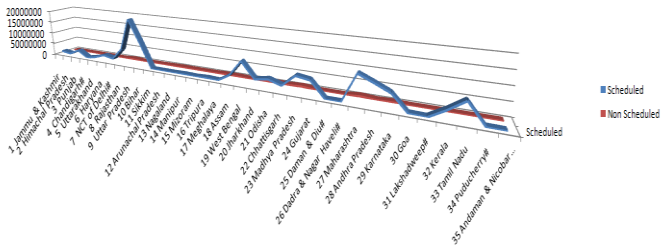


Fig 7: DISTRIBUTION OF POPULATION BY SCHEDULED AND OTHER LANGUAGES

Family	Language	Clusters	Silhouette	1971	1981	1991	2001	2011
Indo-European	Hindi	C1	0.5	202767971	257749009	329518087	422048642	528347193
Indo-European	Bengali	C3	0.694874	44792312	51298319	69595738	83369769	97237669
Tibeto-Burmese	Marathi	C3	0.707401	41765190	49452922	62481681	71936894	83026680
Indo-European	Telugu	C3	0.710824	44756923	50624611	66017615	74002856	81127740
Indo-European	Tamil	C5	0.5	37690106	0	53006368	60793814	69026881
Indo-European	Gujarati	C4	0.629467	25865012	33063267	40673814	46091617	55492554
Dravidian	Urdu	C4	0.615366	28620895	34941435	43406932	51536111	50772631
Indo-European	Kannada	C4	0.690104	21710649	25697146	32753676	37924011	43706512
Indo-European	Odia	C4	0.695299	19863198	23021528	28061313	33017446	37521324
Indo-European	Malayalam	C4	0.695175	21938760	25700705	30377176	33066392	34838819
Dravidian	Punjabi	C4	0.645396	14108443	19611199	23378744	29102477	33124726
Tibeto-Burmese	Assamese	C2	0.674852	8959558	0	13079696	13168484	15311351

Indo-European	Maithili	C2	0.692281	6130026	7522265	7766921	12179122	13583464
Indo-European	Santali	C2	0.723979	3786899	4332511	5216325	6469600	7368192
Indo-European	Kashmiri	C2	0.728175	2495487	3176975	0	5527698	6797587
Indo-European	Nepali	C2	0.734525	1419835	1360636	2076645	2871749	2926168
Indo-European	Sindhi	C2	0.734362	1676875	2044389	2122848	2535485	2772264
Austro-Asiatic	Dogri	C2	0.733495	1299143	1530616	0	2282589	2596767
Tibeto-Burmese	Konkani	C2	0.734544	1508432	1570108	1760607	2489015	2256502
Dravidian	Manipuri	C2	0.733599	791714	901407	1270216	1466705	1761079
Dravidian	Bodo	C2	0.732622	556576	0	1221881	1350478	1482929
Indo-European	Sanskrit	C2	0.728902	2212	6106	49736	14135	24821

Table 4: k-Means Clustered Scheduled Language Data

Family	**Language**	**Clusters**	**Silhouette**	**1971**	**1981**	**1991**	**2001**	**2011**
Austro-Asiatic	Bhumij	C2	0.727241	51651	50384	45302	47443	27506
Austro-Asiatic	Gadaba	C2	0.730672	20420	28027	28158	26262	40976
Austro-Asiatic	Ho	C1	0.670389	751389	783301	949216	1042724	1421418
Austro-Asiatic	Juang	C2	0.730123	12172	19038	16858	23708	30378
Austro-Asiatic	Kharia	C2	0.543868	191421	212605	225556	239608	297614
Austro-Asiatic	Khasi	C1	0.673683	479028	628846	912283	1128575	1431344
Austro-Asiatic	Koda/Kora	C2	0.731329	14333	23113	28200	43030	47268
Austro-Asiatic	Korku	C5	0.661625	307434	347661	466073	574481	727133
Austro-Asiatic	Korwa	C2	0.729184	15097	48079	27485	34586	28453

Austro-Asiatic	Munda	C5	0.665443	309293	377492	413894	469357	505922
Austro-Asiatic	Mundari	C1	0.667493	771253	742739	861378	1061352	1128228
Austro-Asiatic	Nicobarese	C2	0.730517	17971	21542	26261	28784	29099
Austro-Asiatic	Savara	C5	0.542411	222018	209092	273168	252519	409549
Dravidian	Coorgi/Kodagu	C2	0.706191	72085	92678	97011	166187	113857
Dravidian	Gondi	C4	0.62576	1688284	1913262	2124852	2713790	2984453
Dravidian	Jatapu	C2	0.729138	36450	22850	25730	39331	20028
Dravidian	Khond/Kondh	C2	0.648446	196316	195793	220783	118597	155548
Dravidian	Kisan	C2	0.675765	73847	159327	162088	141088	206100
Dravidian	Kolami	C2	0.712526	66868	83690	98281	121855	128451
Dravidian	Konda	C2	0.729934	33720	23258	17864	56262	60699
Dravidian	Koya	C5	0.593029	211877	240245	270994	362070	407423
Dravidian	Kui	C1	0.598421	351017	521585	641662	916222	941488
Dravidian	Kurukh/Oraon	C4	0.614977	1235665	1333670	1426618	1751489	1988350
Dravidian	Malto	C2	0.672358	0	100177	108148	224926	234991
Dravidian	Parji	C2	0.726319	73912	35758	44001	51216	52349
Dravidian	Tulu	C4	0.610037	1158419	1417224	1552259	1722768	1846427
Indo-European	Afghani/Kabuli/Pashto	C2	0.726831	0	0	0	11086	21677
Indo-European	Bhili/Bhilodi	C3	0.5	3399285	4293314	5572308	9582957	10413637
Indo-European	Bishnupuriya	C2	0.726909	0	0	59233	77545	79646
Indo-European	English	C2	0.595379	191595	202440	178598	226449	259678
Indo-European	Halabi	C5	0.598095	346259	534825	534313	593443	766297
Indo-European	Khandeshi	C4	0.487867	251896	1216789	973709	2075258	1860236
Indo-European	Lahnda	C2	0.724246	41935	42879	27386	92234	108791
Indo-European	Shina	C2	0.727556	0	10275	14858	0	34390
Semito-	Arabic/Arbi	C2	0.730759	23318	28116	21975	51728	54947

Hamitic								
Tibeto-Burmese	Adi	C2	0.654727	99228	125371	158409	198462	248834
Tibeto-Burmese	Anal	C2	0.729518	6592	11074	12156	23191	27217
Tibeto-Burmese	Angami	C2	0.712182	43569	41761	97631	132225	152796
Tibeto-Burmese	Ao	C2	0.63353	75381	102321	172449	261387	260008
Tibeto-Burmese	Balti	C2	0.726138	0	40136	48498	0	20053
Tibeto-Burmese	Bhotia	C2	0.703858	33226	31438	55483	81012	229954
Tibeto-Burmese	Chakhesang	C2	0.727782	0	0	30985	11415	19846
Tibeto-Burmese	Chakru/Chokri	C2	0.7262	0	0	48207	83560	91216
Tibeto-Burmese	Chang	C2	0.730342	15816	22442	32478	62408	66852
Tibeto-Burmese	Deori	C2	0.729711	14937	0	17901	27960	32376
Tibeto-Burmese	Dimasa	C2	0.716845	40149	0	88543	111961	137184
Tibeto-Burmese	Gangte	C2	0.72777	6033	0	13695	14500	16542
Tibeto-Burmese	Garo	C1	0.637709	411731	417888	675642	889479	1145323
Tibeto-Burmese	Halam	C2	0.731221	19197	19364	29322	38275	38915
Tibeto-Burmese	Hmar	C2	0.725703	38207	36365	65204	83404	98988
Tibeto-Burmese	Kabui	C2	0.721235	50814	52113	68925	94758	122931
Tibeto-Burmese	Karbi/Mikir	C5	0.6216	199121	12600	366229	419534	528503
Tibeto-Burmese	Khezha	C2	0.730633	11363	16637	13004	40768	41625
Tibeto-Burmese	Khiemnungan	C2	0.730753	14414	17880	23544	37755	61983
Tibeto-	Kinnauri	C2	0.726705	45472	52864	61794	65097	83561

Burmese								
Tibeto-Burmese	Koch	C2	0.731011	14256	16694	26179	31119	36434
Tibeto-Burmese	Kom	C2	0.727905	0	10062	13548	14673	15108
Tibeto-Burmese	Konyak	C2	0.656153	72338	76092	137722	248109	244477
Tibeto-Burmese	Kuki	C2	0.727858	32560	49478	58263	52873	83968
Tibeto-Burmese	Ladakhi	C2	0.722083	0	56738	72274	0	104618
Tibeto-Burmese	Lahauli	C2	0.728625	16749	18728	22027	22646	11574
Tibeto-Burmese	Lakher	C2	0.731145	11867	16091	22947	34751	42429
Tibeto-Burmese	Lalung	C2	0.729831	10650	0	33746	27072	33921
Tibeto-Burmese	Lepcha	C2	0.73054	33360	27814	39342	50629	47331
Tibeto-Burmese	Liangmei	C2	0.729937	0	0	27478	34232	49811
Tibeto-Burmese	Limbu	C2	0.730739	0	20258	28174	37265	40835
Tibeto-Burmese	Lotha	C2	0.70255	36949	58116	85802	170001	179467
Tibeto-Burmese	Lushai/Mizo	C5	0.5949	271554	384528	538842	674756	830846
Tibeto-Burmese	Mao	C2	0.69918	35381	58813	77810	0	240205
Tibeto-Burmese	Maram	C2	0.729117	0	0	10144	37340	32460
Tibeto-Burmese	Maring	C2	0.729317	0	11663	15268	22326	25814
Tibeto-Burmese	Miri/Mishing	C5	0.645231	180684	0	390583	551224	629954
Tibeto-Burmese	Mishmi	C2	0.731119	22354	24182	29000	33955	44100
Tibeto-Burmese	Mogh	C2	0.731004	12378	17458	28135	30639	36665
Tibeto-	Monpa	C2	0.728325	26369	33320	43226	55876	13703

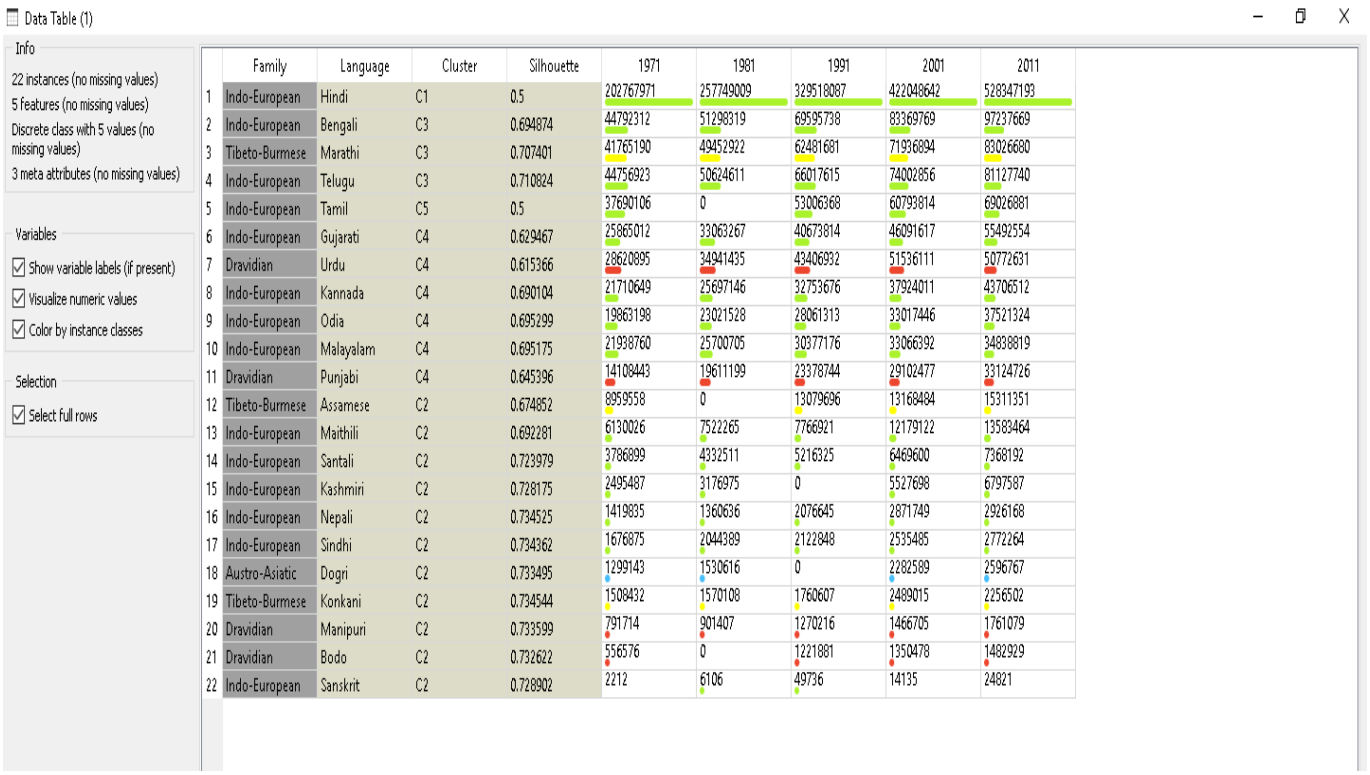
Burmese								
Tibeto-Burmese	Nissi/Dafla	C2	0.55449	114678	140986	173791	211485	406532
Tibeto-Burmese	Nocte	C2	0.730534	25263	23776	30441	32957	30839
Tibeto-Burmese	Paite	C2	0.729202	27157	32607	49237	64100	79507
Tibeto-Burmese	Pawi	C2	0.72994	10560	11656	15346	24965	28639
Tibeto-Burmese	Phom	C2	0.723285	18017	24487	65350	122508	54416
Tibeto-Burmese	Pochury	C2	0.728007	0	0	11231	16744	21654
Tibeto-Burmese	Rabha	C2	0.7032	51146	22351	139365	164770	139986
Tibeto-Burmese	Rai	C2	0.72665	0	0	0	14378	15644
Tibeto-Burmese	Rengma	C2	0.729847	0	15563	37521	61345	65328
Tibeto-Burmese	Sangtam	C2	0.728341	20015	28513	47461	84273	76000
Tibeto-Burmese	Sema	C2	0.705813	65227	95630	166157	103529	10802
Tibeto-Burmese	Sherpa	C2	0.7284	0	14195	16105	18342	16012
Tibeto-Burmese	Tamang	C2	0.727561	0	10059	0	17494	20154
Tibeto-Burmese	Tangkhul	C2	0.701724	58167	79887	101841	142035	187276
Tibeto-Burmese	Tangsa	C2	0.731091	13333	12027	28121	40086	38624
Tibeto-Burmese	Thado	C2	0.684416	51054	57536	107992	190595	229340
Tibeto-Burmese	Tibetan	C2	0.712001	49221	63431	69416	85278	182685
Tibeto-Burmese	Tripuri	C1	0.619681	372579	502067	694940	854023	1011294
Tibeto-Burmese	Vaiphei	C2	0.731262	12209	15618	26185	39673	42748
Tibeto-	Wancho	C2	0.730605	28649	32442	39600	49072	59154

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

Burmese								
Tibeto-Burmese	Yimchungre	C2	0.727277	19609	26672	47227	92144	83259
Tibeto-Burmese	Zeliang	C2	0.729216	0	0	35079	61547	63529
Tibeto-Burmese	Zemi	C2	0.730428	0	10619	22634	34110	50925
Tibeto-Burmese	Zou	C2	0.729325	0	12515	15966	20857	26545

Table 5: k-Means Clustered Non Scheduled Language Data

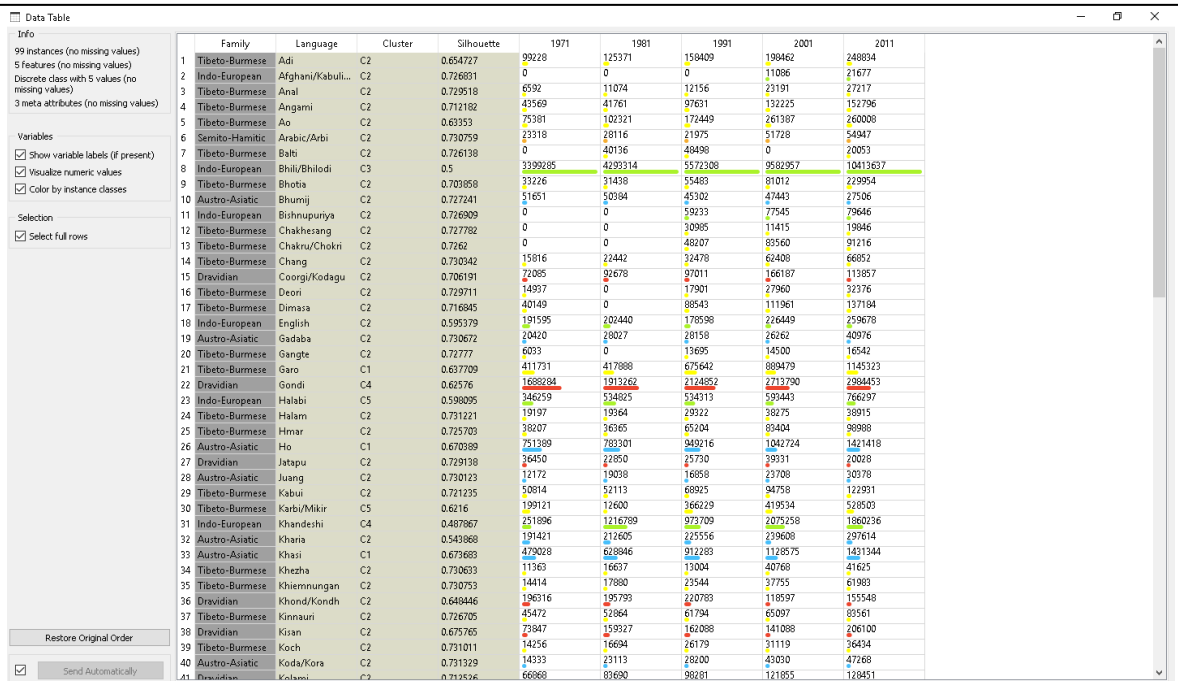


	Family	Language	Cluster	Silhouette	1971	1981	1991	2001	2011
1	Indo-European	Hindi	C1	0.5	202767971	257749009	329518087	422048642	528347193
2	Indo-European	Bengali	C3	0.694874	44792312	51298319	69595738	83369769	97237669
3	Tibeto-Burmese	Marathi	C3	0.707401	41765190	49452922	62481681	71936894	83026680
4	Indo-European	Telugu	C3	0.710824	44756923	50624611	66017615	74002856	81127740
5	Indo-European	Tamil	C5	0.5	37690106	0	53006368	60793814	69026881
6	Indo-European	Gujarati	C4	0.629467	25865012	33063267	40673814	46091617	55492554
7	Dravidian	Urdu	C4	0.615366	28620895	34941435	43406932	51536111	50772631
8	Indo-European	Kannada	C4	0.690104	21710649	25697146	32753676	37924011	43706512
9	Indo-European	Odia	C4	0.695299	19863198	23021528	28061313	33017446	37521324
10	Indo-European	Malayalam	C4	0.695175	21938760	25700705	30377176	33066392	34838819
11	Dravidian	Punjabi	C4	0.645396	14108443	19611199	23378744	29102477	33124726
12	Tibeto-Burmese	Assamese	C2	0.674852	8959558	0	13079696	13168484	15311351
13	Indo-European	Maithili	C2	0.692281	6130026	7522265	7766921	12179122	13583464
14	Indo-European	Santali	C2	0.723979	3786899	4232511	5216325	6469600	7368192
15	Indo-European	Kashmiri	C2	0.728175	2495487	3176975	0	5527698	6797587
16	Indo-European	Nepali	C2	0.734525	1419835	1360636	2076645	2871749	2926168
17	Indo-European	Sindhi	C2	0.734362	1676875	2044389	2122848	2535485	2772264
18	Austro-Asiatic	Dogri	C2	0.733495	1299143	1530616	0	2282589	2596767
19	Tibeto-Burmese	Konkani	C2	0.734544	1508432	1570108	1760607	2489015	2256502
20	Dravidian	Manipuri	C2	0.733599	791714	901407	1270216	1466705	1761079
21	Dravidian	Bodo	C2	0.732622	556576	0	1221881	1350478	1482929
22	Indo-European	Sanskrit	C2	0.728902	2212	6106	49736	14135	24821

Fig 8: Data Table of Scheduled Languages

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021



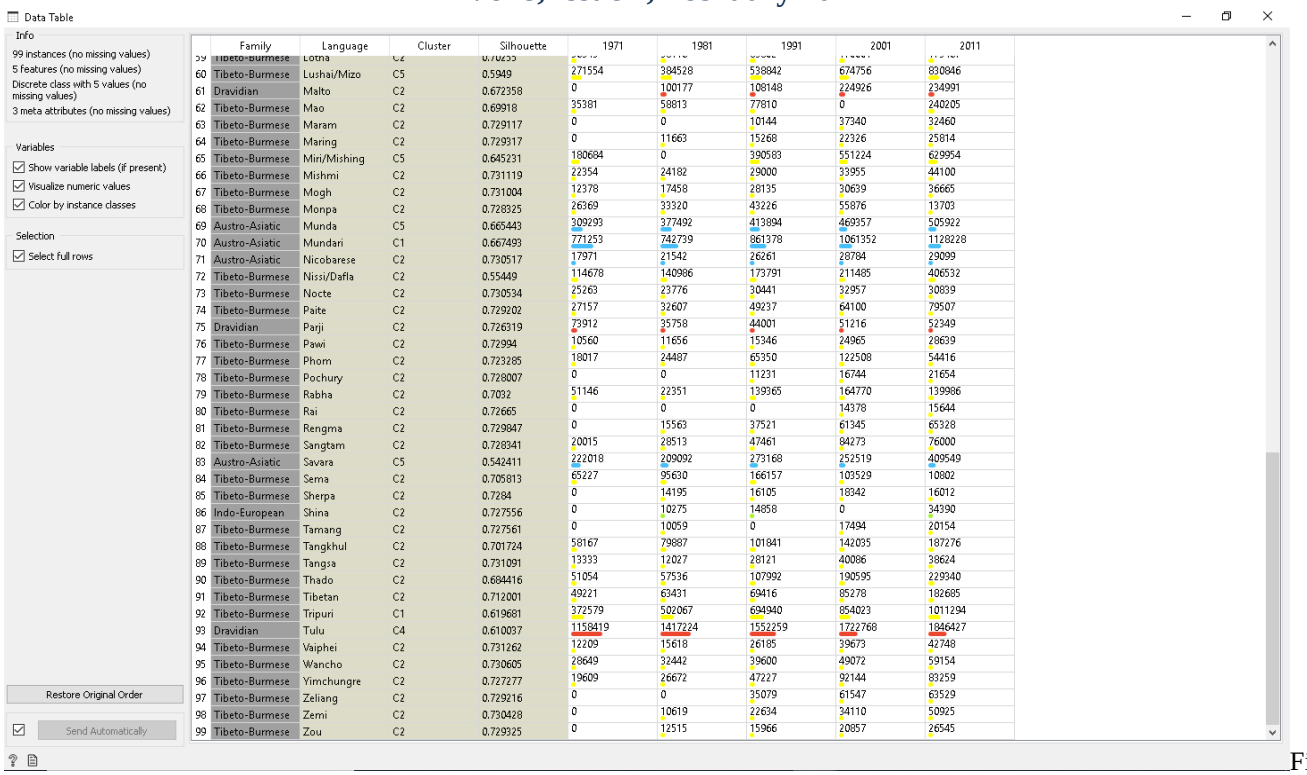
	Family	Language	Cluster	Silhouette	1971	1981	1991	2001	2011
1	Tibeto-Burmese	Adi	C2	0.654727	99228	125371	158409	198462	248834
2	Indo-European	Afghani/Kabuli...	C2	0.726831	0	0	0	11086	21677
3	Tibeto-Burmese	Anal	C2	0.729518	6592	11074	12156	23191	27217
4	Tibeto-Burmese	Angami	C2	0.712182	43569	41761	97631	132225	152796
5	Tibeto-Burmese	Ao	C2	0.63353	75381	102321	172449	261387	260008
6	Semito-Hamitic	Arabic/Arbi	C2	0.730759	23318	28116	21975	51728	54947
7	Tibeto-Burmese	Balti	C2	0.726138	0	40136	48498	0	20053
8	Indo-European	Bhili/Bhilodi	C3	0.5	3399285	4293314	5572308	9582957	10413637
9	Tibeto-Burmese	Bhotia	C2	0.703858	33226	31438	55483	81012	229954
10	Austro-Asiatic	Bhumij	C2	0.727241	51651	50384	45302	47443	27506
11	Indo-European	Bishnupuriya	C2	0.726909	0	0	59233	77545	79646
12	Tibeto-Burmese	Chakhesang	C2	0.727782	0	0	30985	11415	19846
13	Tibeto-Burmese	Chakru/Chokri	C2	0.7262	0	0	48207	83560	91216
14	Tibeto-Burmese	Chang	C2	0.730342	15816	22442	32478	62408	66852
15	Dravidian	Coorgi/Kodagu	C2	0.706191	72085	92678	97011	166187	113857
16	Tibeto-Burmese	Deori	C2	0.729711	14937	0	17901	27960	32376
17	Tibeto-Burmese	Dimasa	C2	0.716845	40149	0	88543	111961	137184
18	Indo-European	English	C2	0.595379	191595	202440	178598	226449	259678
19	Austro-Asiatic	Gadaba	C2	0.730672	20420	28027	28158	26262	40976
20	Tibeto-Burmese	Gangte	C2	0.72777	6033	0	13695	14500	16542
21	Tibeto-Burmese	Garo	C1	0.637709	411731	417888	675642	889479	1145323
22	Dravidian	Gondi	C4	0.62576	1688284	1913262	2124852	2713790	2984453
23	Indo-European	Halabi	C5	0.598095	346259	534825	534313	599443	766297
24	Tibeto-Burmese	Halam	C2	0.731221	19197	19364	29322	38275	38915
25	Tibeto-Burmese	Hmar	C2	0.725703	38207	36365	65204	83404	98988
26	Austro-Asiatic	Ho	C1	0.670389	751389	783301	949216	1042724	1421418
27	Dravidian	Jatapu	C2	0.729138	36450	22850	25730	39331	20028
28	Austro-Asiatic	Juang	C2	0.730123	12172	19038	16858	23708	30378
29	Tibeto-Burmese	Kabui	C2	0.721235	50814	52113	68925	94758	122931
30	Tibeto-Burmese	Karbi/Mikir	C5	0.6216	199121	12600	366229	419534	528503
31	Indo-European	Khandeshi	C4	0.487867	251896	1216789	973709	2075258	1860236
32	Austro-Asiatic	Kharia	C2	0.543868	191421	212605	225556	239608	297614
33	Austro-Asiatic	Khasi	C1	0.673683	479028	628846	912283	1128575	1431344
34	Tibeto-Burmese	Khezha	C2	0.730633	11363	16637	13004	40768	41625
35	Tibeto-Burmese	Khiemnungan	C2	0.730753	14414	17880	23544	37755	61983
36	Dravidian	Khond/Kondh	C2	0.648446	196316	195793	220783	118597	155548
37	Tibeto-Burmese	Kinnauri	C2	0.726705	45472	52864	61794	65097	83561
38	Dravidian	Kisan	C2	0.675765	73847	159327	162088	141088	206100
39	Tibeto-Burmese	Koch	C2	0.731011	14256	16694	26179	31119	36434
40	Austro-Asiatic	Koda/Kora	C2	0.731329	14333	23113	28200	43030	47268
41	Dravidian	Kolami	C2	0.712526	66868	83690	98281	121855	128451

	Family	Language	Cluster	Silhouette	1971	1981	1991	2001	2011
37	Tibeto-Burmese	Kinnauri	C2	0.726705	45472	52864	61794	65097	83561
38	Dravidian	Kisan	C2	0.675765	73847	159327	162088	141088	206100
39	Tibeto-Burmese	Koch	C2	0.731011	14256	16694	26179	31119	36434
40	Austro-Asiatic	Koda/Kora	C2	0.731329	14333	23113	28200	43030	47268
41	Dravidian	Kolami	C2	0.712526	66868	83690	98281	121855	128451
42	Tibeto-Burmese	Kom	C2	0.727905	0	10062	13548	14673	15108
43	Dravidian	Konda	C2	0.729934	33720	23258	17864	56262	60699
44	Tibeto-Burmese	Konyak	C2	0.656153	72338	76092	137722	248109	244477
45	Austro-Asiatic	Korku	C5	0.661625	307434	347661	466073	574481	727133
46	Austro-Asiatic	Korwa	C2	0.729184	15097	48079	27485	34586	28453
47	Dravidian	Koya	C5	0.593029	211877	240245	270994	362070	407423
48	Dravidian	Kui	C1	0.598421	351017	521585	641662	916222	941488
49	Tibeto-Burmese	Kuki	C2	0.727858	32560	49478	58263	52873	83968
50	Dravidian	Kurukh/Oraon	C4	0.614977	1225665	1333670	1426618	1751489	1988350
51	Tibeto-Burmese	Ladakhi	C2	0.722083	0	56738	72274	0	104618
52	Tibeto-Burmese	Lahauli	C2	0.728625	16749	18728	22027	22646	11574
53	Indo-European	Lahnda	C2	0.724246	41935	42879	27386	92234	108791
54	Tibeto-Burmese	Lakher	C2	0.731145	11867	16091	22947	34751	42429
55	Tibeto-Burmese	Lalung	C2	0.729831	10650	0	33746	27072	33921
56	Tibeto-Burmese	Lepcha	C2	0.73054	33360	27814	39342	50629	47331
57	Tibeto-Burmese	Liangmei	C2	0.729937	0	0	27478	34232	49811
58	Tibeto-Burmese	Limbu	C2	0.730739	0	20258	28174	37265	40835
59	Tibeto-Burmese	Lotha	C2	0.70255	36949	58116	85802	170001	179467
60	Tibeto-Burmese	Lushai/Mizo	C5	0.5949	271554	384528	538842	674756	830846
61	Dravidian	Malto	C2	0.672358	0	100177	108148	224926	234991
62	Tibeto-Burmese	Mao	C2	0.69918	35381	58813	77810	0	240205
63	Tibeto-Burmese	Maram	C2	0.729117	0	0	10144	37340	32460
64	Tibeto-Burmese	Maring	C2	0.729317	0	11663	15268	22326	25814
65	Tibeto-Burmese	Miri/Mishing	C5	0.645231	180684	0	390583	551224	629954
66	Tibeto-Burmese	Mishmi	C2	0.731119	22354	24182	29000	33955	44100
67	Tibeto-Burmese	Mogh	C2	0.731004	12378	17458	28135	30639	36665
68	Tibeto-Burmese	Monpa	C2	0.728325	26369	33320	43226	55876	13703
69	Austro-Asiatic	Munda	C5	0.665443	309293	377492	413894	469357	505922
70	Austro-Asiatic	Mundari	C1	0.667493	771253	742739	861378	1061352	1128228
71	Austro-Asiatic	Nicobarese	C2	0.730517	17971	21542	26261	28784	29099
72	Tibeto-Burmese	Nissi/Dafla	C2	0.55449	114678	140986	173791	211485	406532
73	Tibeto-Burmese	Nocte	C2	0.730534	25263	23776	30441	32957	30839
74	Tibeto-Burmese	Paite	C2	0.729202	27157	32607	49237	64100	79507
75	Dravidian	Parji	C2	0.726319	23912	35758	44001	51216	52349
76	Tibeto-Burmese	Pawi	C2	0.72994	10560	11656	15346	24965	28639
77	Tibeto-Burmese	Phom	C2	0.723285	18017	24487	65350	122508	54416

Fig 10: Data Table of Non Scheduled Languages

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021



	Family	Language	Cluster	Silhouette	1971	1981	1991	2001	2011
60	Tibeto-Burmese	Lushai/Mizo	C5	0.5949	271554	384528	538842	674756	830846
61	Dravidian	Malto	C2	0.672358	0	100177	108148	224926	234991
62	Tibeto-Burmese	Mao	C2	0.69918	35381	58813	77810	0	240205
63	Tibeto-Burmese	Maram	C2	0.729117	0	0	10144	37340	32460
64	Tibeto-Burmese	Maring	C2	0.729317	0	11663	15268	22326	25814
65	Tibeto-Burmese	Miri/Mishing	C5	0.645231	180684	0	390583	551224	629954
66	Tibeto-Burmese	Mishmi	C2	0.731119	22354	24182	29000	33955	44100
67	Tibeto-Burmese	Mogh	C2	0.731004	12378	17458	28135	30639	36665
68	Tibeto-Burmese	Monpa	C2	0.728325	26369	33320	43226	55876	13703
69	Austro-Asiatic	Munda	C5	0.665443	309293	377492	413894	469357	505922
70	Austro-Asiatic	Mundari	C1	0.667493	771253	742739	861378	1061352	1128228
71	Austro-Asiatic	Nicobarese	C2	0.730517	17971	21542	26261	28784	29099
72	Tibeto-Burmese	Nissi/Dafla	C2	0.55449	114678	140986	173791	211485	406532
73	Tibeto-Burmese	Nocte	C2	0.730534	25263	23776	30441	32957	30839
74	Tibeto-Burmese	Paite	C2	0.729202	27157	32607	49237	64100	79507
75	Dravidian	Parji	C2	0.726319	73912	35758	44001	51216	52349
76	Tibeto-Burmese	Pawi	C2	0.72994	10560	11656	15346	24965	28639
77	Tibeto-Burmese	Phom	C2	0.723285	18017	24487	65350	122508	54416
78	Tibeto-Burmese	Pochury	C2	0.728007	0	0	11231	16744	21654
79	Tibeto-Burmese	Rabha	C2	0.7032	51146	22351	139365	164770	139986
80	Tibeto-Burmese	Rai	C2	0.72665	0	0	0	14378	15644
81	Tibeto-Burmese	Rengma	C2	0.729847	0	15563	37521	61345	65328
82	Tibeto-Burmese	Sangtam	C2	0.728341	20015	28513	47461	84273	76000
83	Austro-Asiatic	Savara	C5	0.542411	222018	209092	273168	252519	409549
84	Tibeto-Burmese	Sema	C2	0.705813	65227	95630	166157	109529	10802
85	Tibeto-Burmese	Sherpa	C2	0.7284	0	14195	16105	18342	16012
86	Indo-European	Shina	C2	0.727556	0	10275	14858	0	34390
87	Tibeto-Burmese	Tamang	C2	0.727561	0	10059	0	17494	20154
88	Tibeto-Burmese	Tangkhul	C2	0.701724	58167	79887	101841	142035	187276
89	Tibeto-Burmese	Tangsa	C2	0.731091	13333	12027	28121	40086	38624
90	Tibeto-Burmese	Thado	C2	0.684416	51054	57536	107992	190595	229340
91	Tibeto-Burmese	Tibetan	C2	0.712001	49221	63431	69416	85278	182685
92	Tibeto-Burmese	Tripuri	C1	0.619681	372579	502067	694940	854023	1011294
93	Dravidian	Tulu	C4	0.610037	1158419	1417224	1552259	1722768	1846427
94	Tibeto-Burmese	Vaiphei	C2	0.731262	12209	15618	26185	39673	42748
95	Tibeto-Burmese	Wancho	C2	0.730605	28649	32442	39600	49072	59154
96	Tibeto-Burmese	Yimchungre	C2	0.727277	19609	26672	47227	92144	83259
97	Tibeto-Burmese	Zeliang	C2	0.729216	0	0	35079	61547	63529
98	Tibeto-Burmese	Zemi	C2	0.730428	0	10619	22634	34110	50925
99	Tibeto-Burmese	Zou	C2	0.729325	0	12515	15966	20857	26545

Fig 11: Data Table of Non Scheduled Languages

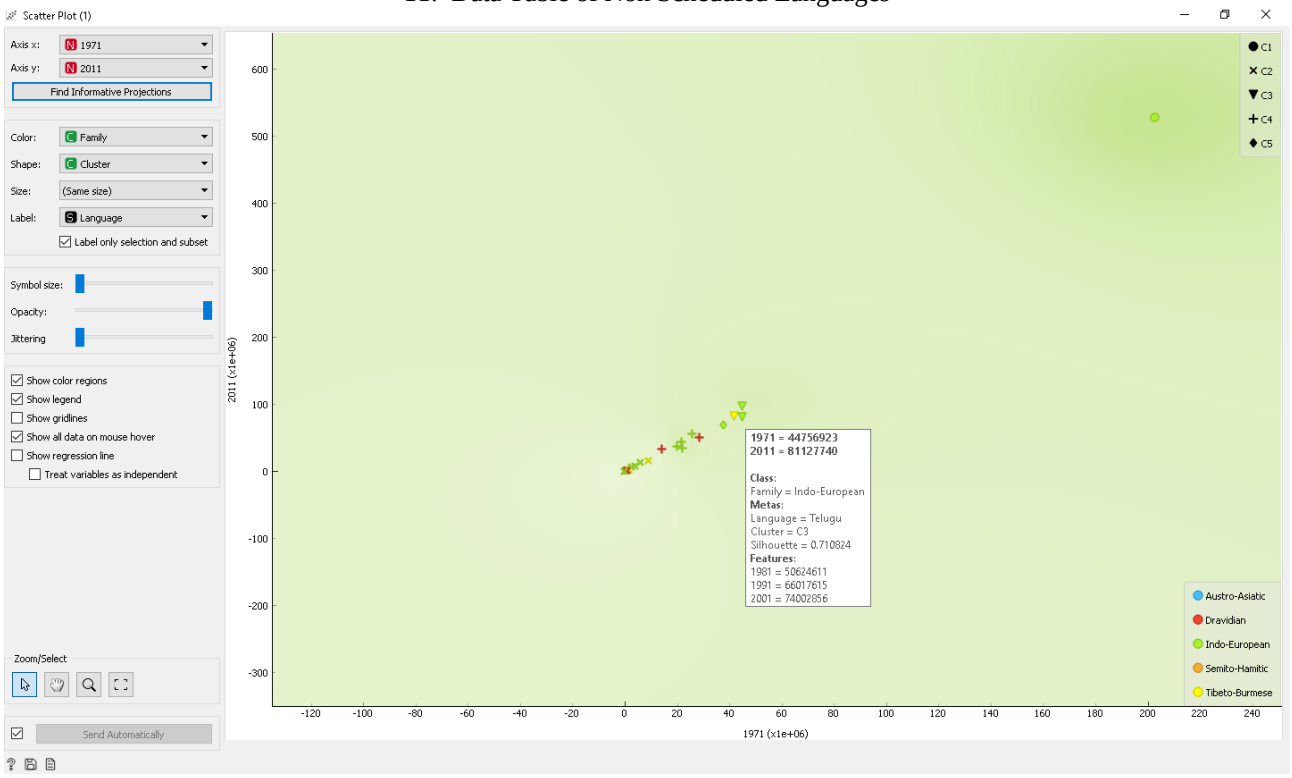


Fig 12: Scheduled language using Scatter Plot from k-Means

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

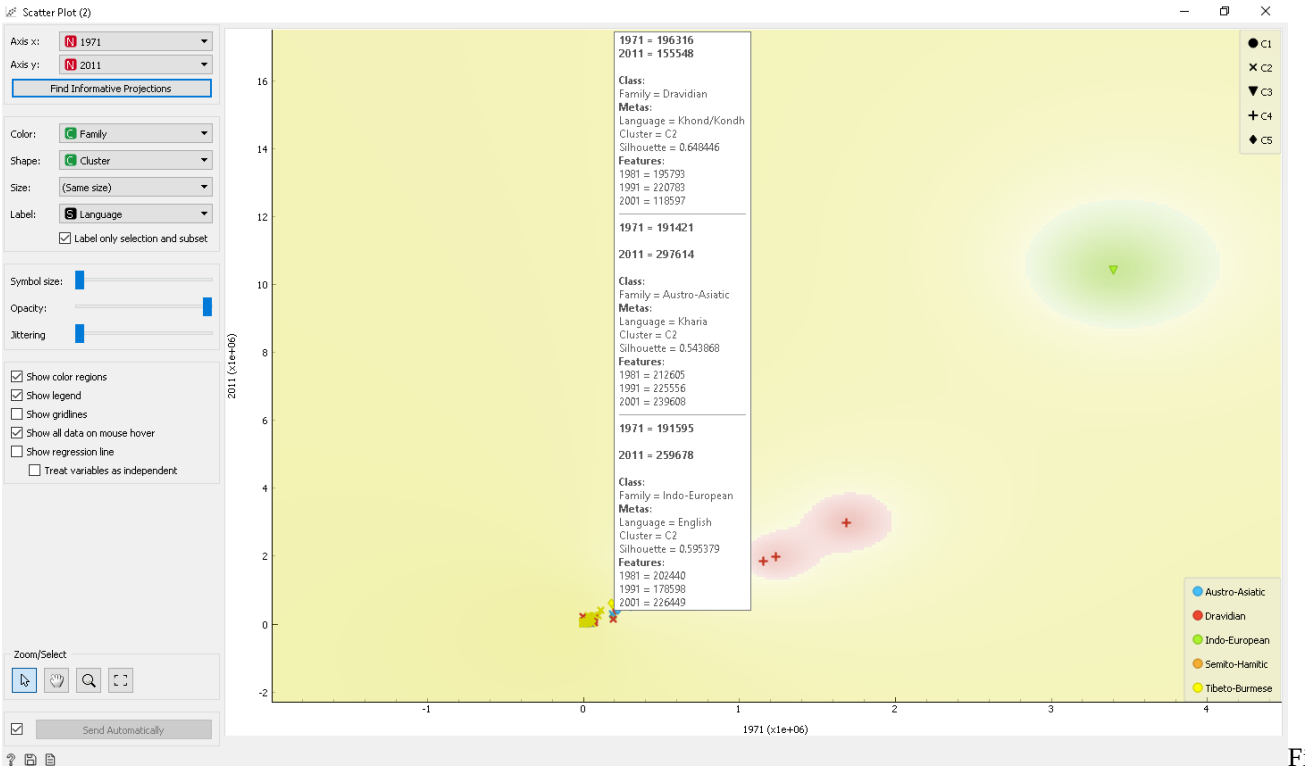


Fig 13: Non schedule language using Scatter plot from k-Means

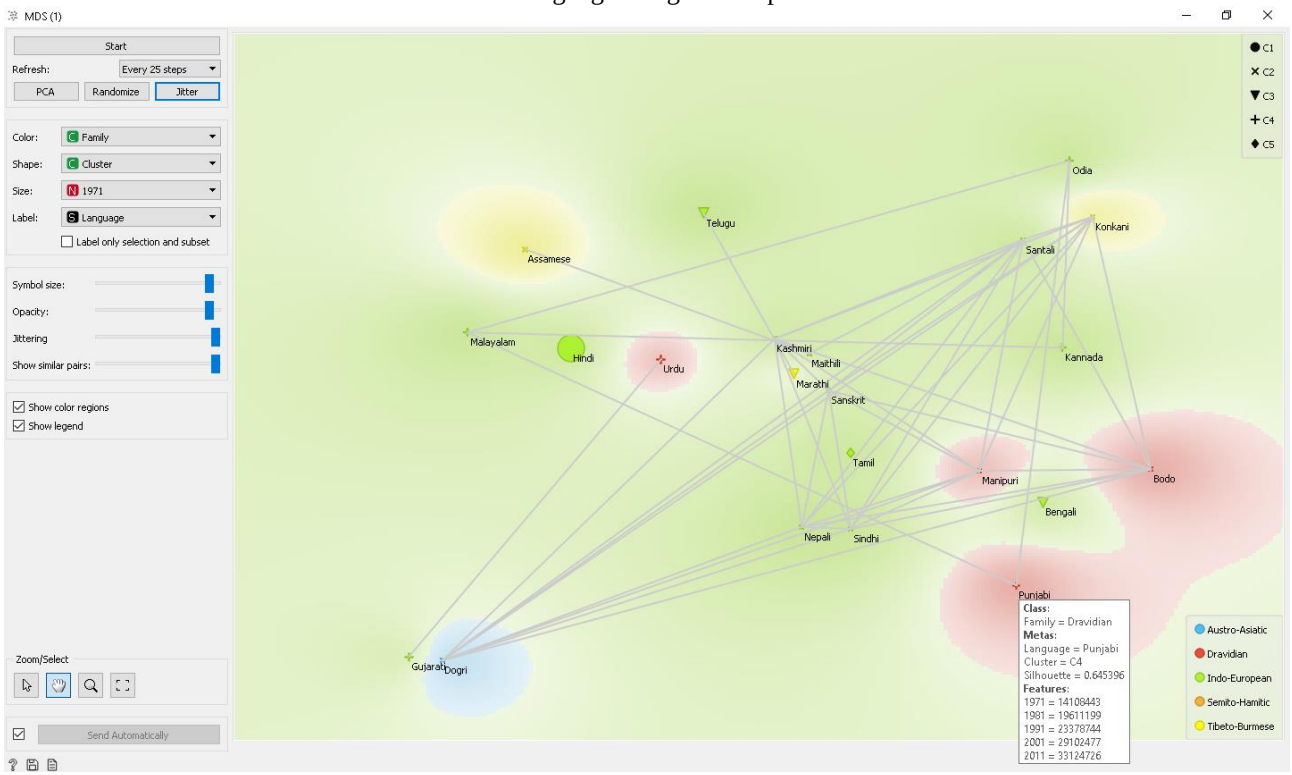


Fig 14: Scheduled language using MDS from k-Means

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

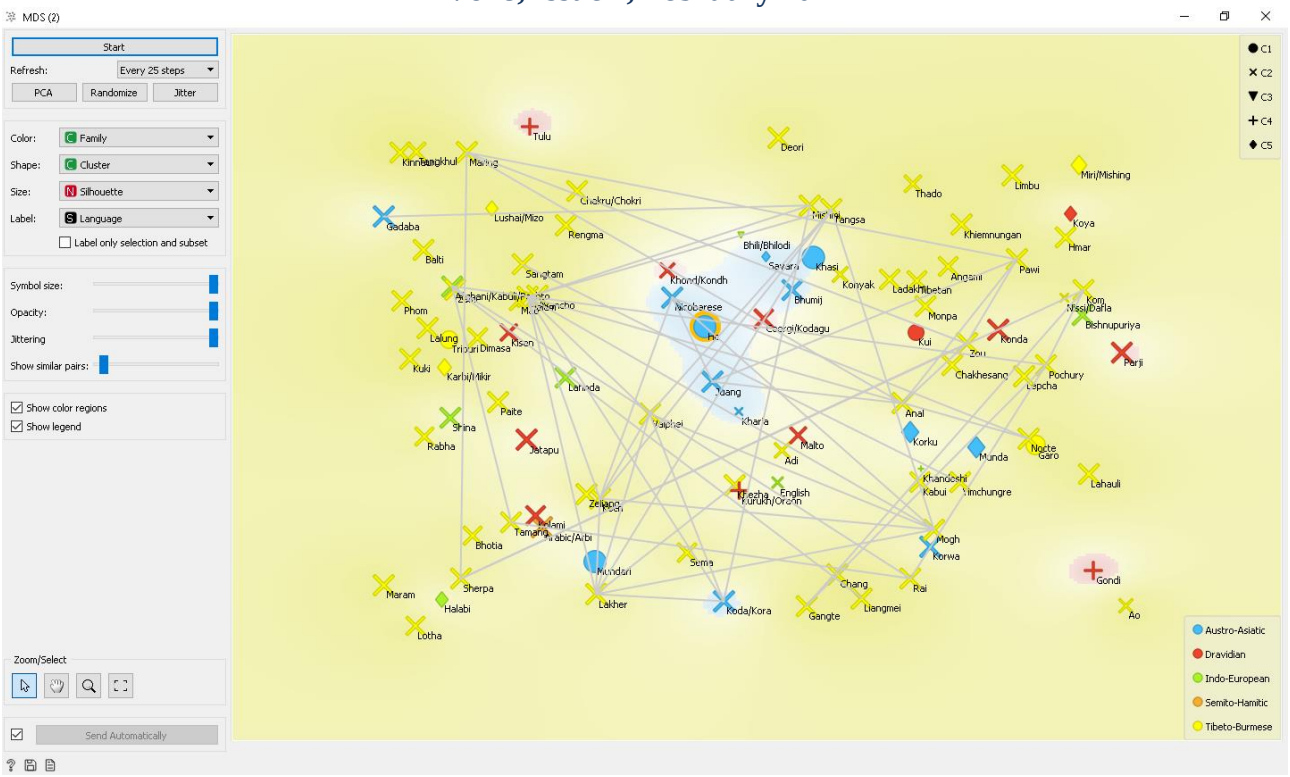


Fig 15: Non Scheduled language using MDS from k-Means

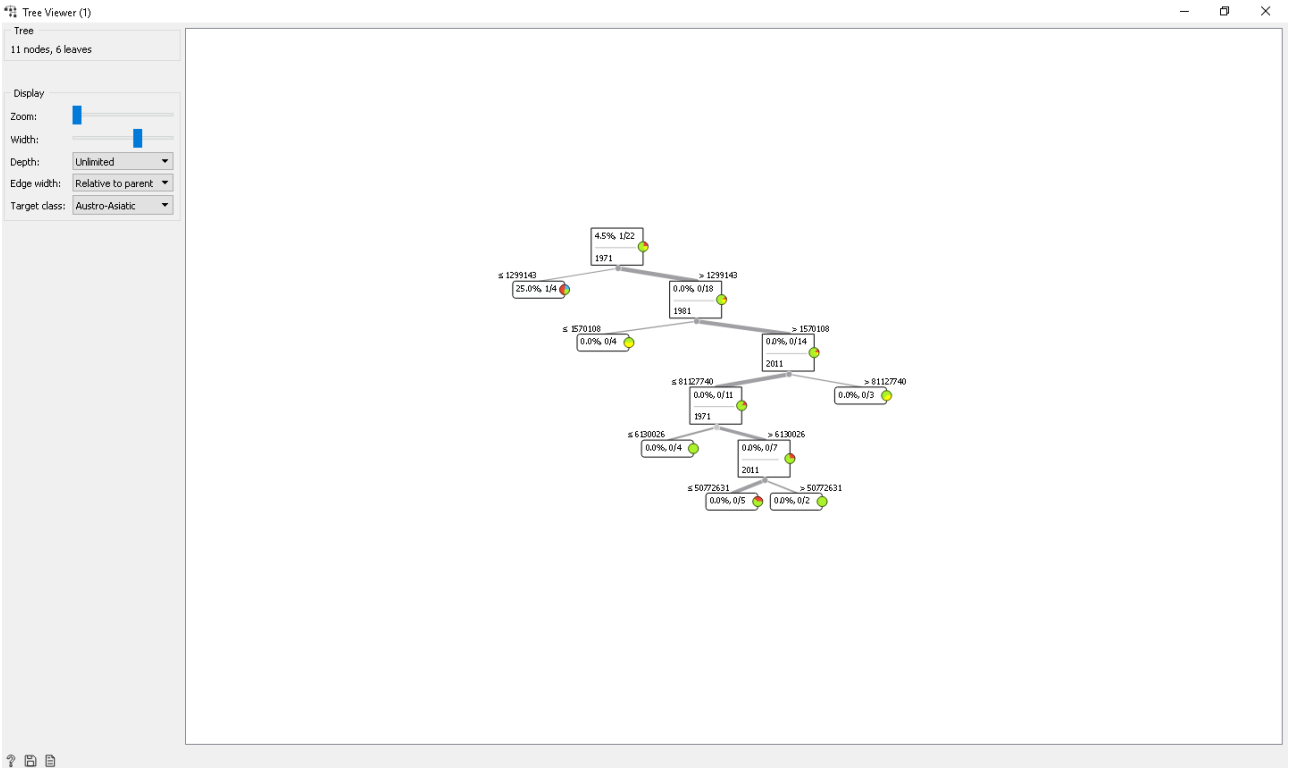


Fig 16: Scheduled language family of Austro-Asiatic

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

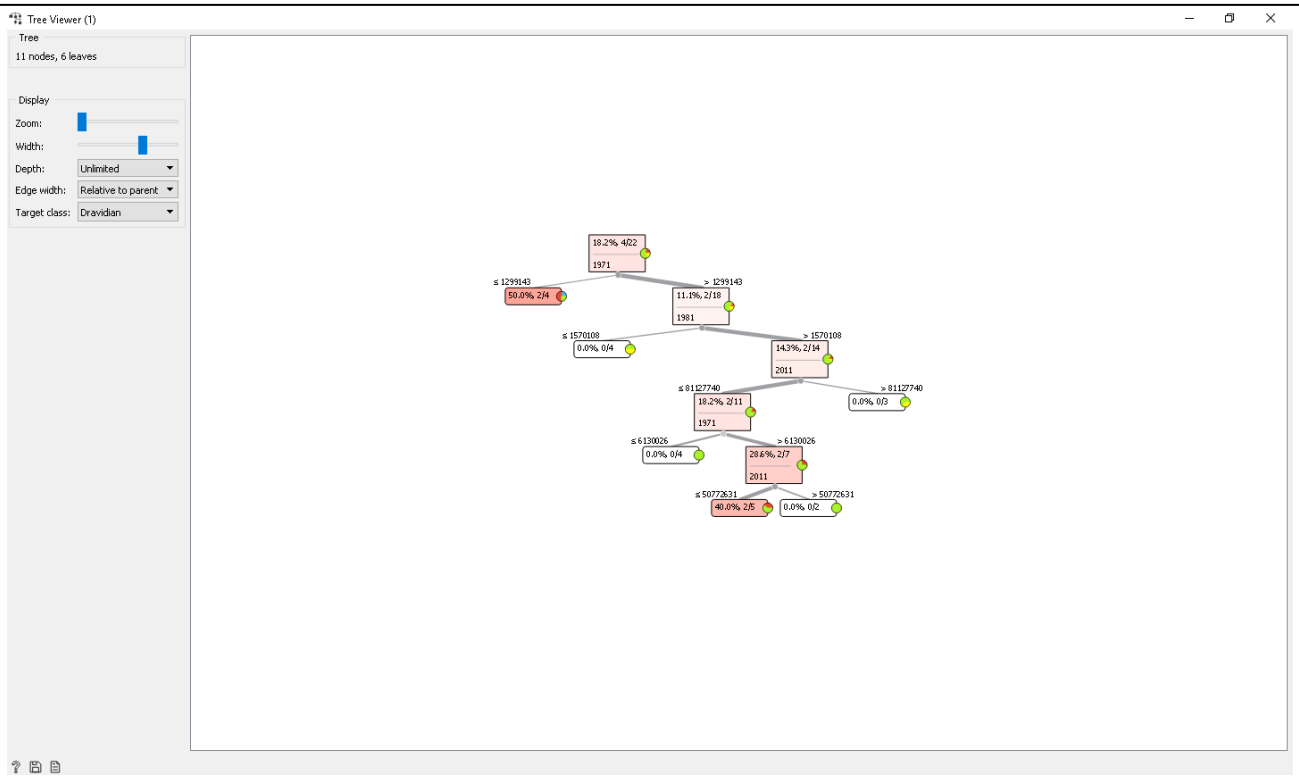


Fig 17: Scheduled language family of Dravidian

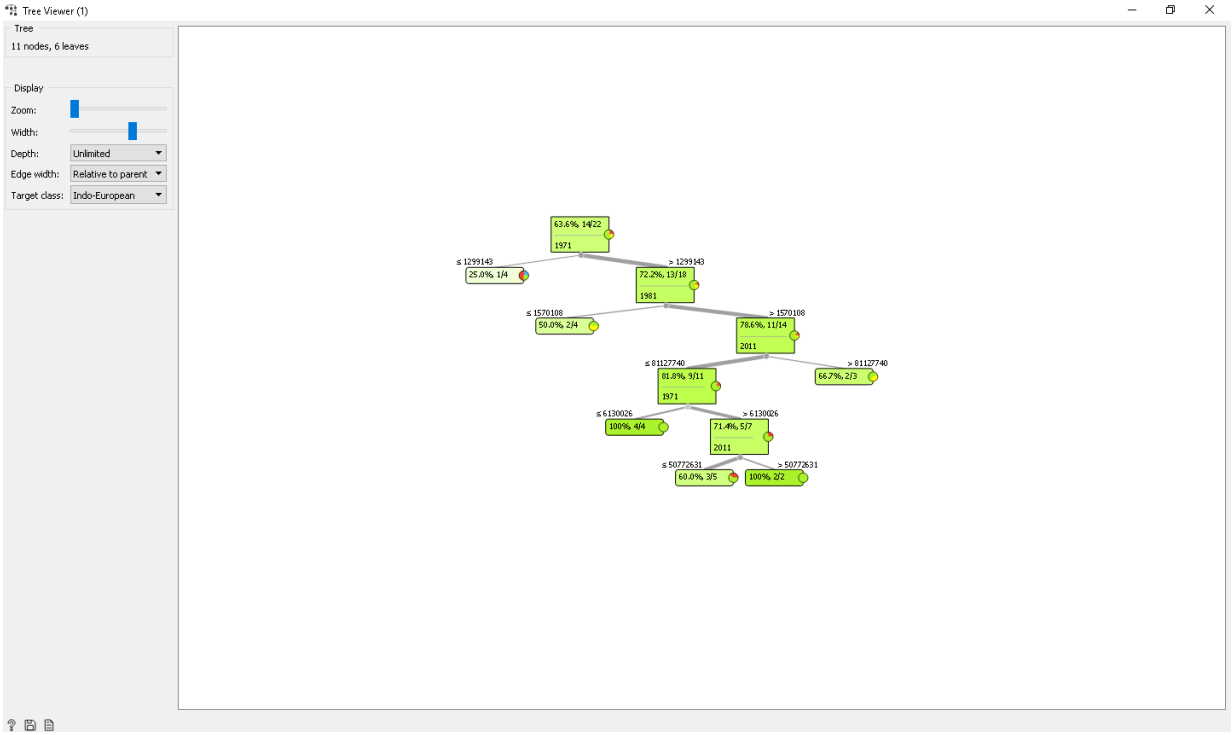


Fig. 18: Scheduled language family of Indo-European

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

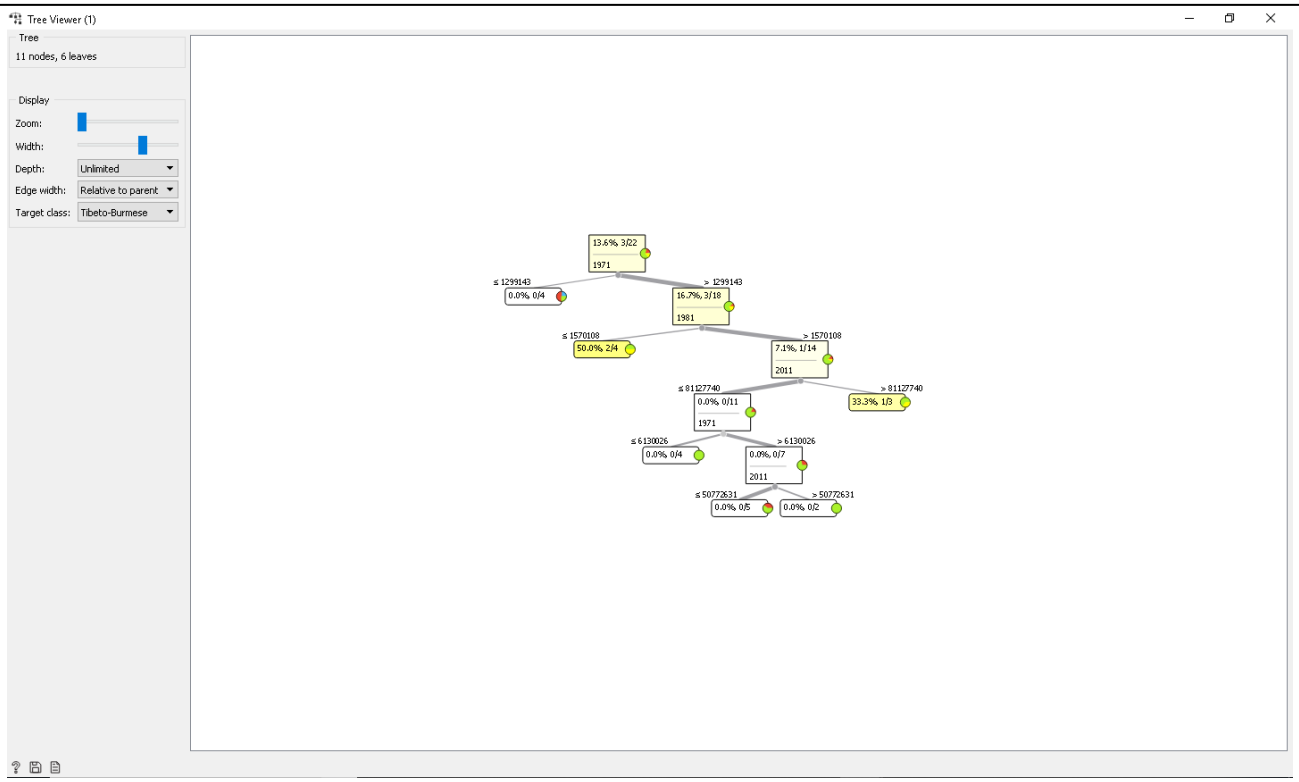


Fig. 19: Scheduled language family of Tibeto-Burmese

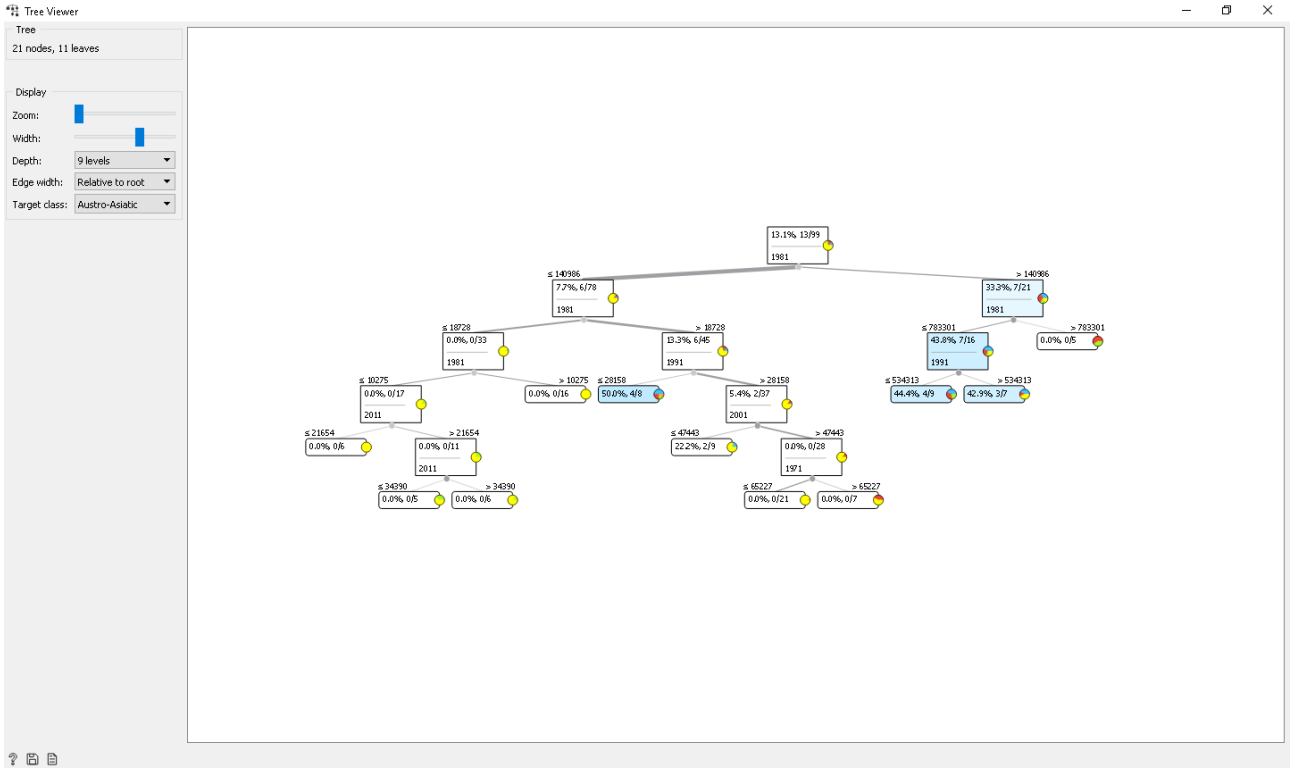


Fig. 20: Treeview analysis of Non Scheduled Austro-Asiatic language family

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

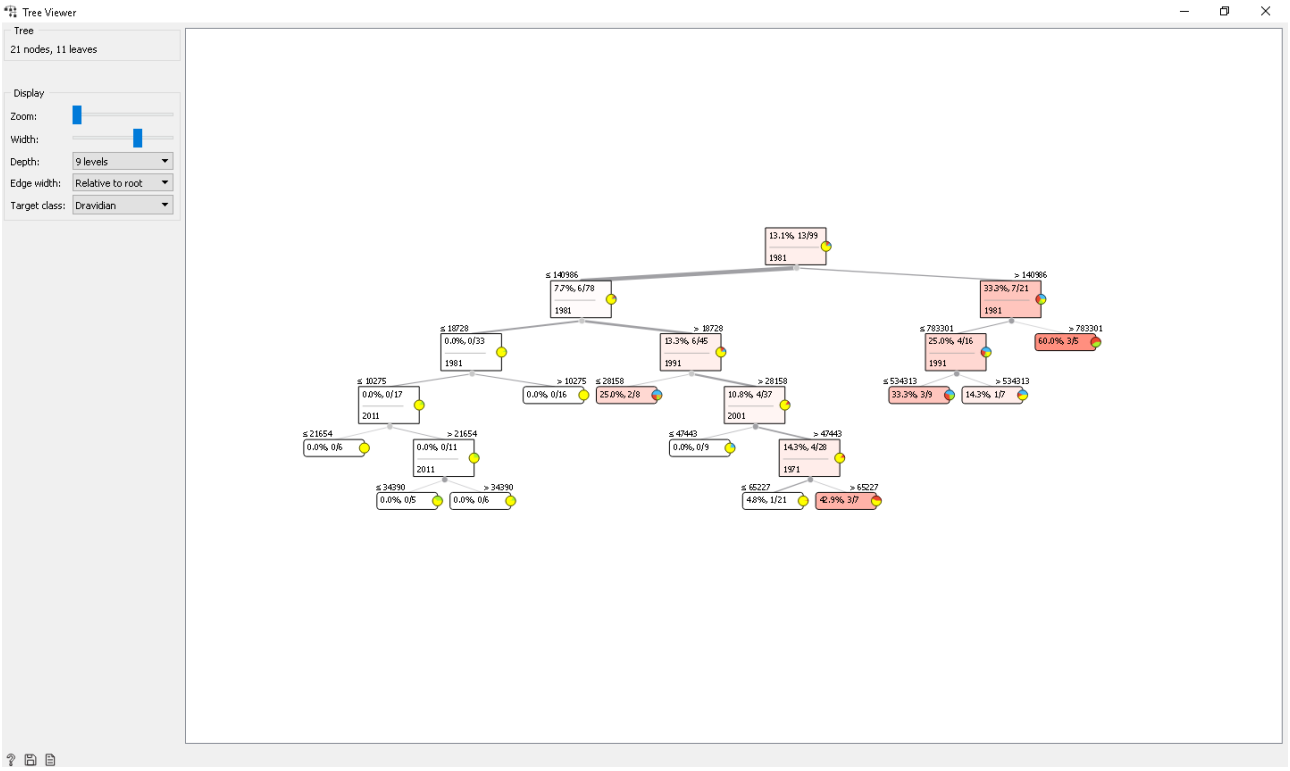


Fig. 21: Treeview analysis of Non Scheduled Dravidian language family

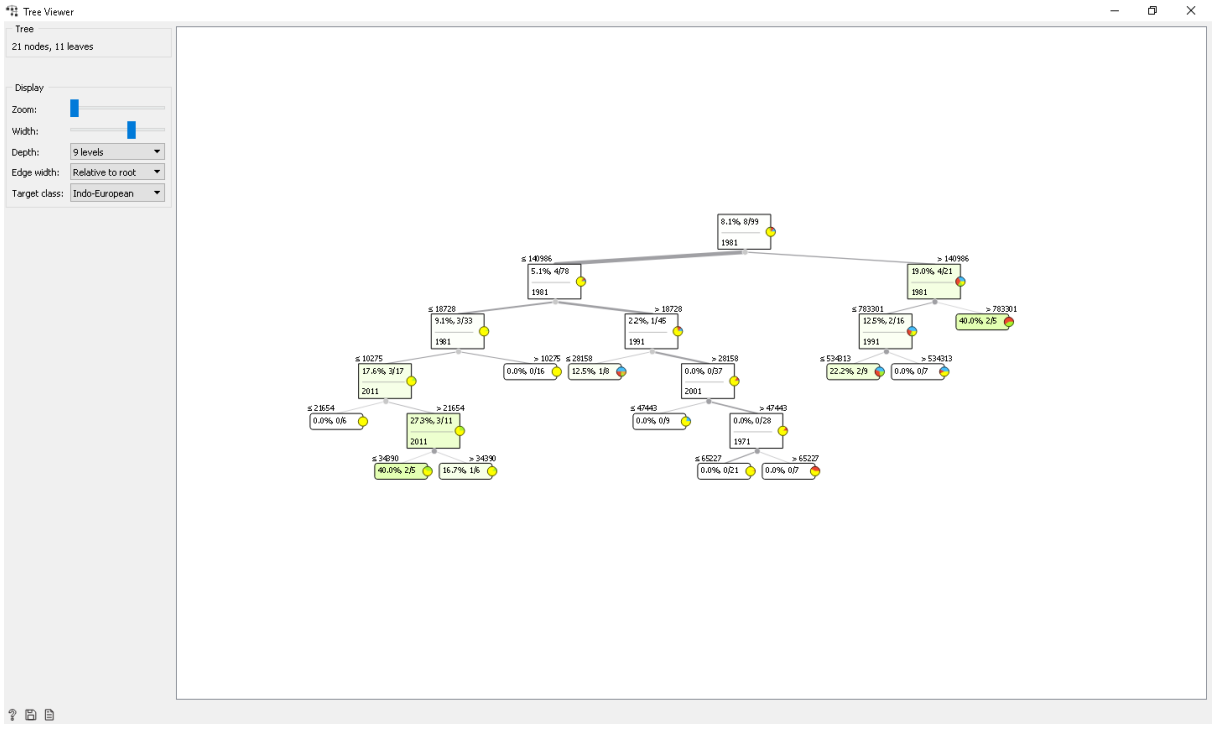


Fig. 22: Treeview analysis of Non Scheduled Indo-European language family

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

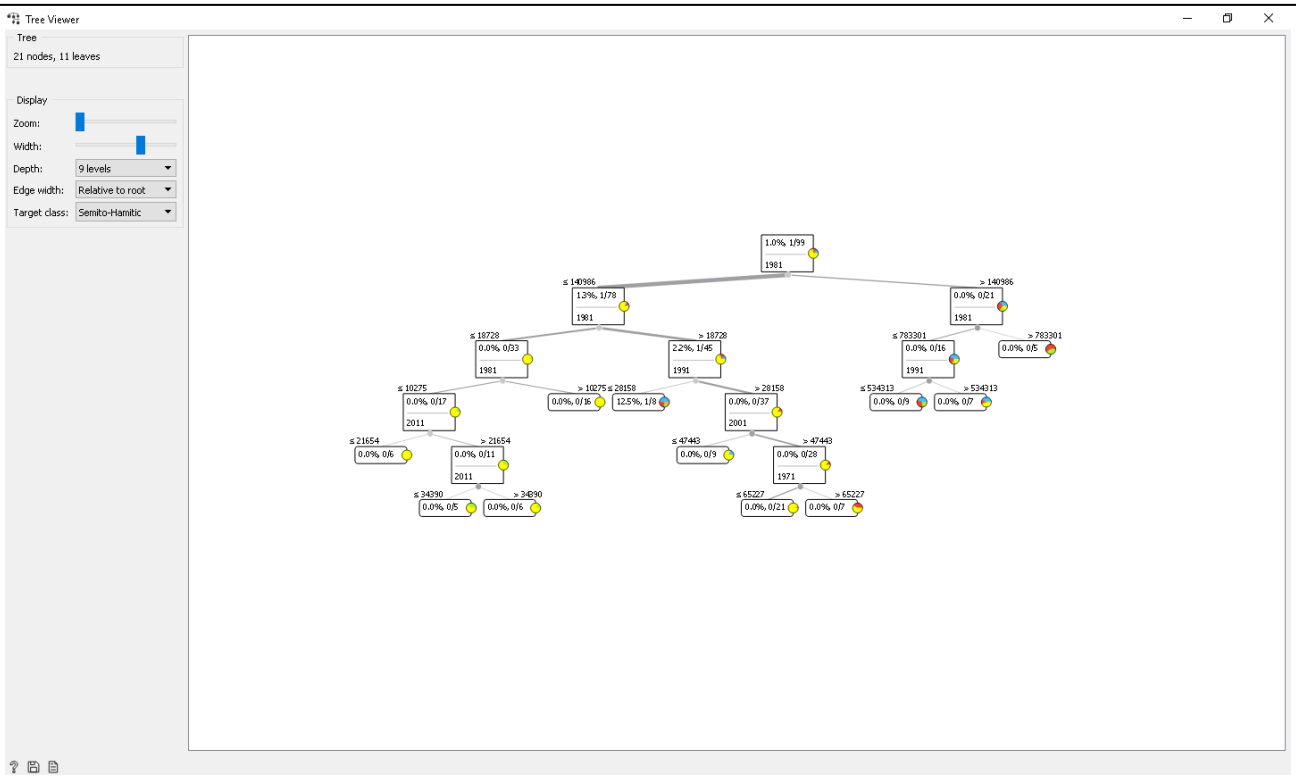


Fig. 23: Treeview analysis of Non Scheduled Semito Hamitic language family

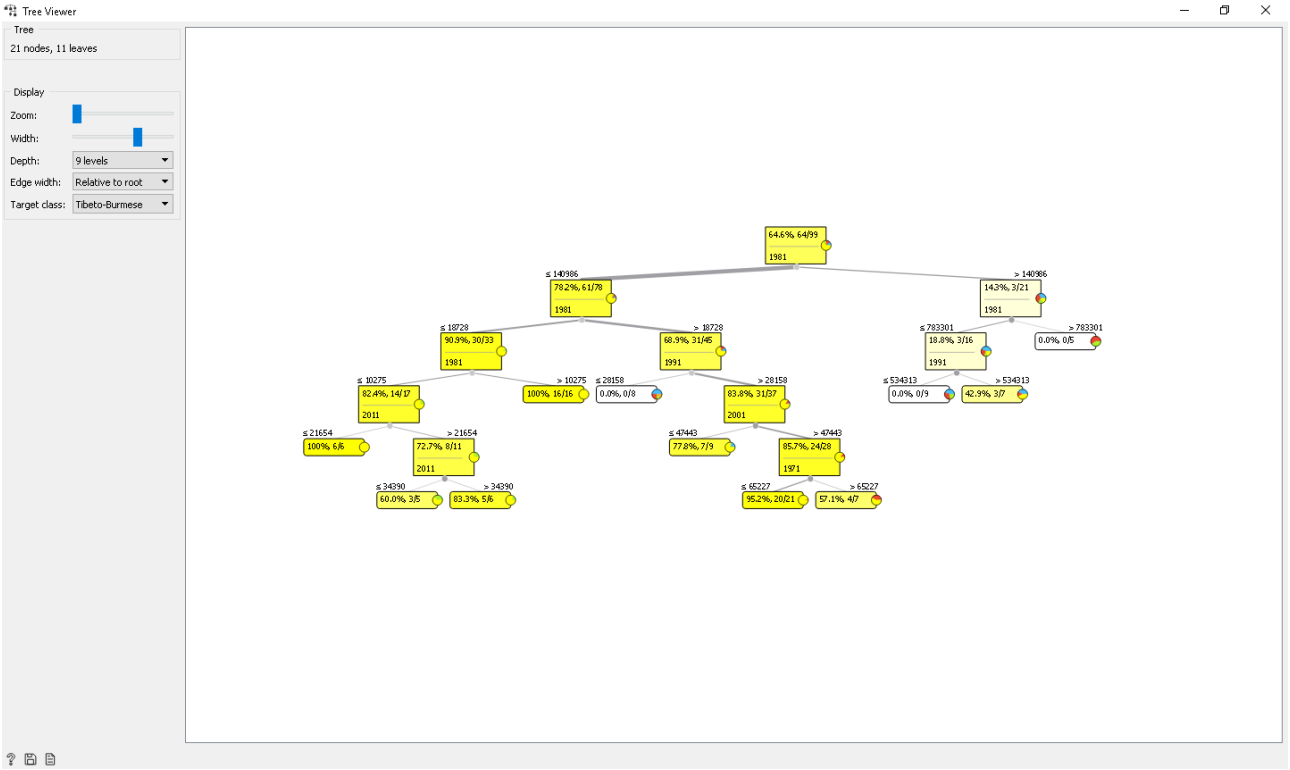


Fig 24: Treeview analysis of Non Scheduled Tibeto-Burmese language family

V. RESULTS

The k-Means is unsupervised algorithm to analyze groups based on Indian families based on silhouette values and easily classified groups with strength and weakness. K-Means clustering the special group of clusters based on silhouette score and random initialization of iteration. From Fig.2 observing that 'Hindi' Language of Color occupies more colors visually. In scheduled language 'Hindi' speakers are more company to any other languages. Family groups are arranging color size as the family group between range 1 and 5 in Fig. 3. Similarly Non Scheduled languages are occupy more color visually appears in Bhilli, Gondi, Kurukh, etc. (Fig.4) and size of the color arranging indicate that the family group between range 1 to 5 in Fig 5. The comparison between scheduled and non-scheduled language in family group wise based on visualization structure and it is very easily identified the strength and weakness of the family in Fig 6. State wise calculations of scheduled and non scheduled languages are explored in Fig 7. So, from these easily identified the strength of the scheduled and nonscheduled in various region of the language development. Uttarpradesh, Maharashtra, Bihar, etc are highest speakers of scheduled languages and Madhyapradesh, Maharashtra, Rajasthan, etc are the highest speakers of nonscheduled languages (Fig. 7). Data table are representing visually by color and the spread of the color is more indicates that the languages is used more number of people. Data table is easily identified the strength and weakness of the language based on visually representation. Data table of orange mining is mainly feature of visualize numerical values are very important factor to visualize the data compare to any software. From Fig. 8 we observed that visualize color indicates the occupies the value and easily find out the more number of language speakers (Hindi, Bengali, Telugu) and less language speakers (Sanskrit, Bodo, Manipuri) from scheduled language. The family wise also visually observes the strength and weakness based on the color. The colors are indicates to family group of individual color. Similarly, Fig. 9, Fig.10, Fig.11 are the Data Table of non scheduled languages. In non scheduled language of Bhili, Gondi, Kurukh, etc are easily visualizing the data in family. The color of Indo-European, Dravidian, etc are represents the strength and weakness in visually. The family groups are 5 with 5 different colors. Scheduled languages are classified using k-Means with following clusters.

C1={ Hindi},
C2={Assamese, Maithlli, Santali, Kashmiri, Nepali, Sindhi, Dogri, Konkani, Manipuri, Bodo, Sanskrit},
C3={Bengali, Marathi, Telugu},
C4={Gujarati, Urdu, Kannada, Odia, Malayalam, Punjabi}
C5={Tamil}

From Fig. 12 are representing the scheduled language using scatter plot from k-means. Scatter plot provides the X and Y axis based on year wise (1971, 2011), family field represents the color, cluster field represent the shape, and language field represent the label. In cluster C1, Hindi is more speakers compare to any type languages and it also represents the Indo-European family strength is more visually by color. Some of the Dravidian (Manipuri, Bodo), Tibeto-Burmese (Assamese, Konkani), Austro-Asiatic (Dogri), Indo-European (Maithilli, Santali, Kashmiri, Nepali, Sindhi, Sanskrit) are belongs to cluster C2 and having less speakers among family groups. The pair between the languages groups based on the family in visual representation using scatter plot. Similarly (Fig. 13) from the non-scheduled languages k-Means clustering is used the year wise variations with multiple family groups are explored in visually with all descriptions. Family groups are representing the color of the family. Clustering groups are differentiated with cover regions and different styles with each cluster group and additional option for labeling any of the field. Language speakers are more number of Indo-European (bhilli), Dravidian (gondi, kurukh) from non scheduled language using cluster and the cluster C5 belongs to more number of Tibeto-Burmese languages, Indo-European (Afghani, Shina, Bishnupriya, Lahnda), Dravidian (Kolami, Kisan, Malto), Semito-Hamitic (Arabic) are grouped with each other with less speaker languages. Similarly Scheduled languages and Non Scheduled languages are described in another multi dimensional scaling plot (MDS) and also visually explored the strength and weakness of the pair language with family group. Family group is represented in color group and PCA, Randomize are graphical representation to view pair of groups visually with multiple clusters among color regions. Language family wise comparison of strength and weakness are easily identified by the Fig. 16 to Fig. 24 for both scheduled languages and non scheduled languages. Austro-Asiatic is less number of language group in scheduled language and Semito-Hamitic is less number of language group in non scheduled languages. The family groups are explored the range of values between language group based on year wise and the large family groups in Indo-European from scheduled languages with large family groups Tibeto-burmese from non scheduled groups visually explored.

VI. CONCLUSION

Visual method analysis is prospect way to describe any kind of data. The language structure depends on the culture, tradition, etc. Languages speech communication is essentially to distribute the information is properly. So, the information communication technologies visually explore the data in understandable method. Visual methods are sharing the knowledge elaborately with big data. The mathematical, statistical analysis is required to predict many aspects of the reason in many problems and through this to identify many solutions. Visual analysis methods are exploring the characteristics of the languages strength and weakness. In future, Data can be expanded to diversity of languages, Mapping of regional language using ArcGIS, How languages are developed in our country and mother tongue development of each state. The people strength is to find multi languages developing with many aspects.

Acknowledgments

The authors thank Director, Bharathiar University, and Coimbatore for his encouragement in undertaking this study for R and D Development.

References

[1] Hammarström, Harald; Forkel, Robert; Haspelmath, Martin, eds. (2017). "Indo-Aryan". Glottolog 3.0. Jena, Germany: Max Planck Institute for the Science of Human History.

[2] Gurpreet Singh Josan & Jagroop Kaur (2011)' Punjabi to Hindi Statistical Machine Translaiteration', International Journal of Information Technology and Knowledge Management July-December 2011, Volume 4, No. 2, pp. 459-463.

[3] Black A.W., Zen H., and Tokuda K., "Statisticalparametric speech synthesis," in Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing, Honolulu, USA, 2007.

[4] S. Chandra Pammi and Prahallad K., "POS tagging and chunkingusing decisionforests," in Proceedingsof Workshop on Shallow Parsing in South Asian Languages, IJCAI, Hyderabad, India, 2007.

[5] PrahalladL., PrahalladK., and GanapathirajuM., "A simple approach for building transliterationeditorsfor indian languages," Journal of Zhejiang University Science, vol. 6A, no. 11, pp. 1354–1361,2005.

[6] GanapathirajuM., BalakrishnanM., BalakrishnanN., and Reddy R., "Om: One tool for many (Indian) languages," Journal of Zhejiang University Science, vol. 6A, no. 11, pp. 1348–1353,2005.

[7] Vijaya M. S., Shivapratap G., Soman K. P (2010), 'English to Tamil Transliteration using One Class Support Vector Machine', International Journal of Applied Engineering Research, Volume 5, Number 4, 641-652.

[8] Rohit Gupta, Pulkit Goyal and Sapan Diwakar (2010), 'Transliteration among Indian Languages using WX Notation by. Semantic Approaches in Natural Language Processing', Proceedings of the Conference on Natural Language Processing 2010. Pages 147-151, Universaar, Universitätsverlag des Saarlandes Saarland University Press, Presses universitaires de la Sarre.

[9] Antony P J, Ajith V P and Soman K P (2010), 'Statistical Method for English to Kannada Transliteration', Lecturer Notes in Computer ScienceCommunications in Computer and Information Science (LNCS-CCIS), Volume 70, 356-362, DOI: 10.1007/978-3-642-12214-9_57.

[10] Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nded.

[11] Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.

[12] Aounallah, M., Quirion, S., Mineau, G.W.: Distributed Data Mining vs. Sampling Techniques: A Comparison. In: Tawfik, A.Y., Goodwin, S.D. (eds.) Canadian AI 2004. LNCS (LNAI), vol. 3060, pp. 454–460. Springer, Heidelberg (2004).

[13] Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann (2006).

[14] Pyle, D.: Data Preparation for Data Mining. Morgan Kaufmann (1999)

Online Sources

[15] Census of India 2011 paper 1 of 2018 Language [Internet]. Available: http://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf