ISSN:0975-9646

Risk Computation and Identification in Passport Data Analysis

Sucheta Gulia [1],Dr. Rajan Vohra [2]

[1]*Student of M.Tech(Computer Science),*
[2]*Head of Department(Computer Science & Information Technology)*
P.D.M College of Engineering, Sector 3A, Sarai Aurangabad, Bhadurgarh, Haryana, India

Abstract:**This paper presents a conceptual information for calculation of score and identification of risk in the passport data to find out the fraud entries in the dataset. This paper also gives information about different level of checks applied on the passport dataset to find out the suspicious entries in the dataset. After computation of the score for each entry in the database risk is identified on the basis of low, medium and high according to the set ranges. Dataset taken for this study is the primary database of the passport holders and applican. In this 3 types of datasets are used to find the score. Different database are required for performing different types of checks for risk computation and its identification.**

Key Words: ***Risk Computation, Passport Data Analysis.***

1. INTRODUCTION

The Indian passport is the primary travel document issued by the Government of India to its citizens. It enables the bearer to travel internationally and serves as proof of Indian citizenship as per the "The Passports Act" (1967). The Consular Passport & Visa (CPV) Division of the Ministry of External Affairs, functioning as the central passport organization, is responsible for issuance of Indian passports on demand to all eligible Indian citizens. Passports are issued at 37 locations across the country and at 162 Indian missions (High Commissions, Embassies and Consular posts) abroad.

A passport is a government-issued travel document that certifies the identity and nationality of its holder for the purpose of international travel. The elements of identity contained in all standardized passports include information about the holder, including name, date of birth, sex and place of birth.

A passport displays nationality, but not the place of residence of the passport holder. The passport holder is normally entitled to re-enter the country that issued the passport in accordance with the laws of that country, and in some instances of gaining a new citizenship, to enter that country for the first time. A passport does not necessarily grant the passport holder entry into any other country, nor to consular protection while abroad or other privileges, such as immunity from arrest or prosecution. Those rights and privileges, if and when applicable, arise from international treaties.

Risk is the probability or threat of quantifiable damage, injury, liability, loss, or any other negative occurrence that is caused by external or internal vulnerabilities, and that may be avoided through preemptive action. Risk is the potential of losing something of value, weighed against the potential to gain something of value. Values (such as physical health, social status, emotional well being or financial wealth) can be gained or lost when taking risk resulting from a given action, activity and/or inaction, foreseen or unforeseen. Risk can also be defined as the intentional interaction with uncertainty. Risk perception is the subjective judgment people make about the severity of a risk, and may vary person to person. Any human endeavor carries some risk, but some are much riskier than others. In this work risk is calculated with the help of Index Check, Prior Approval Check and Police Verification Check then the total score was calculated and then classify it as High, Medium and Low risk.

1.1. Significance of the problem -

The paper is used to find the fraud instances or entries from the passport database. If we take this prototype for police verification, for index check and for prior approval check then the proposed work will find out the fraud persons which are willing for another passport. From the first part of the problem we will list out all the persons which are holding more than one passport, which is illegal. In this we will calculate the score with the help of databases.

Today most of the people apply for the passport and some of them are fraud applicants i.e. they have the passport but they want more then one these types of activities by the people are not relevant and then the passport officers needs to check all the information about the applicant with the help of different checks and assign score to the applicants and then classify them into different risk types. Computation of score and identification of risk is the main significance of this work.

2. RESEARCH METHODOLOGY

For solving the problem some research techniques and methodologies are used for obtaining the desired result. Some tools and algorithms are required for obtaining the result. Those methods and tools are discussed under this section. Main steps under the research methodologies are:-

1. **Review of literature and research papers**
   First of all literatures and research papers were reviewed for getting more information about the problem and knowing which type of work was done by others on this topic and by which method.
2. **Organize field visits**
   Field visit was organized to NIC (NATIONAL INFORMATICS CENTRE) New Delhi, INDIA, nodal agency for passport preparation. From there we get

information about the flow of work regarding passport databases and related compilation.

3. **Study database attributes and data structure**
   Attributes and Structure of the database was thoroughly studied for finding out useful attributes from the passport. For critical attributes(attriutes that define risk) used in the database first and last page of the passport was studied.
4. **Study the work flow**
   The work flow in the work centres like NIC and TCS (TATA CONSULTANCY SERVICE, SOFTWARE SOLUTION PROVIDER) New Delhi, INDIA which are the main centres of the passport was studied for getting information regarding to the different checks required for getting fraud entries.
5. **Organize the database**
   At last database was organized with useful attributes and populate it.

Data collection and system study -- The research strategy adopted was to conduct a survey to gather information required to solve the problems specified in the work. The fieldwork was conducted at the sites during the period from February 2014 to May 2014. The main data collection techniques used in this research study were semi-structured interviews, group discussion, participant observation, documentation analysis and questionnaires.Visit to NIC (National Information Centre) was conducted a few times for the purpose of the system study required in analysis and collection of required data.

Collection of data – the whole data regarding to passport was mainly collected from the TCS office which now a days handles all the information related to passports. The database consist of 478 entries which is the primary data using random sampelling.

3. RESEARCH BACKGROUND

The concept of score calculation is also applied on the banking sector, for crime profiling, for identifying maximum revenue by the network provider. The score calculation model given by the banking system called the cibit model.

Fraud detection can be supervised or unsupervised. Supervised methods use a database of known fraudulent/legitimate cases from which to construct a model which yields a suspicion score for new cases. Traditional statistical classification methods[1][2] (Hand, 1981; McLachlan, 1992), such as linear discriminant analysis and logistic discrimination, have proved to be effective tools for many applications, but more powerful tools [3,4,5,] (Ripley, 1996; Hand, 1997; Webb, 1999), especially neural networks, have also been extensively applied. Rule-based methods are supervised learning algorithms that produce classifiers using rules of the form If {certain conditions}, Then {a consequent}. Examples of such algorithms include BAYES (Clark and Niblett, 1989), FOIL [7](Quinlan, 1990) and RIPPER [9] (Cohen, 1995). Tree-based algorithms such as CART [10] (Breiman, Friedman, Olshen and Stone, 1984) and C4.5 [8] (Quinlan, 1993) produce classifiers of a similar form. Combinations of some or all of these algorithms can be created using meta-learning algorithms to improve prediction in fraud detection [11] (e.g., Chan, Fan, Prodromidis and Stolfo, 1999).

Traditionally, marketers must first identify customer cluster using a mathematical mode and then implement an efficient campaign plan to target profitable customers (12-15). Useful information is often overlooked, and the potential benefits of increased computational and data gathering capabilities are only partially realized. Only through data mining techniques, it is possible to extract useful pattern and association from the customer data (16).

4. CONCEPTUAL FRAMEWORK

The research objectives of the paper are achieved by the following steps:

1. Collect information about passport holders and applicants then mine this database for getting necessary information only.
2. By using different checks compute the score for each applicant.
3. Then identify the risk and classify the applicants with different types of risk.

These checks are described below :-

a. INDEX CHECK :

Attributes used in the index check database are

1. Social security number of the applicant
2. Name and father's name of the passport holder or applicant
3. Date of birth of the person and address
4. Old passport number if another passport exist in past
5. Index Check flag which gives information whether the applicant cleared this check or not.

From the index check database we can get information about a person that whether a person has another passport or not. If applicant has another passport and it is not expired then another passport was not issued to that person then the not clear flag was assigned to that person. There are different conditions from where a certain index check flag was assigned to a particular applicant, these are –

- When a person has another passport and was not expired then Not Clear Flag was raised for that person.
- When a person does not give its old passport number but (s) he has another passport and the entries in the main database reflected the old passport number of that person then the not clear flag was raised for that person.
- When a person has another old passport but (s) he give wrong passport number, then the not clear flag was raised.
- When a person has another old passport but that passport was expired then the clear flag was raised.
- When a person doesn't have another passport and (s) he applied for fresh passport then clear flag was raised.

The score assigned to Not Clear Flag was 2 and to clear flag was 0.

b. PRIOR APPROVAL CHECK :

Attributes used in the prior approval check database are same as in the Index Check database except the old passport number and Index Check Flag. This database gives information about the persons which are banned or

black listed means which are not applicable for issuing the passport for any of the reason. The score assigned to that person was 8. Prior Approval Check gives information about the applicant, whether (s) he is banned in the country or not?

c.POLICE VERIFICATION CHECK :

Attributes used in the Police Verification check database are same as in the Prior. Two additional attributes are added to database which are residing duration and police verification flag. Minimum residing duration at particular place was 24 months that is 2 years if this condition was not satisfied then the passport was not issued to that person because (s) he did not clear the police verification check. In this database the information about every person was stored. From this database police verifies the person that (s) he gives the correct information or not. If the information is found to be incorrect and residing duration was less than 24 months then the flag assigned to that person was "adverse" which gives 4 score to that person and if all information was found correct then the flag assigned to that person was "clear" and score assigned to that person was 0. Then from all three databases score for each entry was filled and total score was calculated by adding all three scores.

4.1. SCORE CALCULATION AND RISK CLASSIFICATION :

For calculating score we can firstly select the attribute and after that frame the database. According to the database entries in all three databases assign the score to each entry in the passport database. for this following steps are performed :-

4.1.1. **Calculating total score -** total score was calculated by summing the index check score, prior approval check score and police verification score. Then we get total score for every entry in the passport database. Then total score is computed simply getting sum of index check score, PAC score and PV score.

$$S(T) = S(PAC) + S(IC) + S(PVC)$$

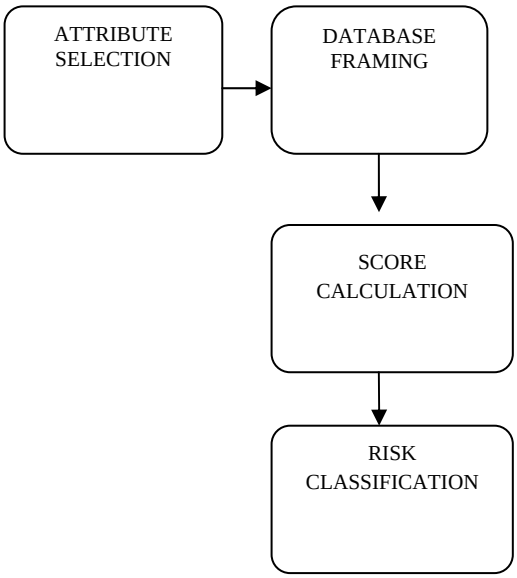


Fig1. Score calculation and risk classification

4.1.2. Identifying Risk Type – Then according to score we will simply classify the score into low, medium and high risk. The range was decided for classifying the risk is as follow –

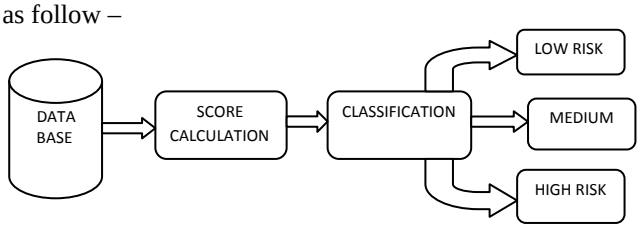


Fig2. Risk computations and classification

- SCORE<4 == LOW RISK,
- 4<=SCORE<6== MEDIUM RISK,
- SCORE>=6 == HIGH RISK

5. RESULT AND DISCUSSION

PREPARING THE DATABASE [17] - For obtaining the result, a database containing total number of passports instances that is 478 was created which is the primary data using random sampling. Passport attributes are selected from passport that are- passport number, file number, name of the applicant, date of birth of the applicant, fathers name, address, place of issue and a unique social security number for each entry. In this problem four databases was used that are- passport database, database for index check, database for prior approval check and database for police verification.

Figure 3 shows the attributes of passport database showing file number, passport number, social security number which is unique for every person, name of the person, fathers name and many more attributes which are critical to thr risk. From those datasets we can calculate score based on different criteria's and assign score according to the instances entries in different datasets. Then according to the score we can classify the risk as low, medium and high.

ATTRIBUTES OF PASSPORT DATABASE

FILE NO.	PASSPORT NO.	SOCIAL SECURITY NO.	NAME	FATHER NAME
FNOO1	PP10010	ABC5715726	SAPNA GIRDHAR JHURANI	THANWARDAS JHAMUMAL CHEKKANI
FN002	PP10011	PQR1234567	RITA CHOWDHRI	DEVINDER KUMAR JAIN
FN003	PP10012	STU8756942	QAMARUDDIN	RIAZUDDIN
FN004	PP10013	DEF2356489	BIJENDER CHANDELIYA	RAJ KUMAR
FN005	PP10014	TUV0854860	MAYUR WADHWA	SURESH CHANDRA WADHWA
FN006	PP10015	UVW0212520	GAGANDEEP SINGH	RAMINER SINGH
FN007	PP10016	XYZ1534009	DUKHTARI BEGUM	NIZAM ALI KHAN
FN008	PP10017	KLM1450547	SARVESH KUMAR SINGH	RAM PRATAPNSINGH
FN009	PP10018	QRS0450261	GOPALAKRISHNA PILLAI	KRISHNA KURUP
FN010	PP10019	FGH1661024	SEEMA RANI	RAM SNEHI
FN011	PP10020	YZA1532201	NEELSHA BHERWANI	SHIV KUMAR BHAMBHANI
FN012	PP10021	DEF2108719	SEEMA RANI	DINESH MOHAN BANSAL
FN013	PP10022	JKL1567652	NAVEEN CHAUDHRY	RAM LAL CHAUDHARY
FN014	PP10023	MNO1571738	SANJIV KUMAR BAHADUR	AJOY KUMAR BAHADUR
FN015	PP10024	BCD1664202	AVNISH KUMAR GUPTA	RAM AVTAR GUPTA
FN016	PP10025	HIJ2534215	JAGPAL SINGH	HUKAM SINGH

Fig3.Attributes of the Passport Database

DATABASE FOR PRIOR APPROVAL CHECK

SOCIAL SECURITY NO.	NAME	FATHERS NAME	DOB	ADDRESS	STATE
FGH1661024	SEEMA RANI	RAM SNEHI	17-Jun-65	QAZI HOZ	DELHI
QAW0579263	RAJINDER SINGH	MEHAR SINGH	01-Mar-63	CHANDER VIHAR NEAR NILOTHI	DELHI
KCS0611344	Bindu Chhikara	RAMESH KUMAR	27-Oct-64	MODI NAGAR	GUJRAT
RSH0047994	Praveen Dhankher	SANJEEV KUMAR	25-Aug-76	SIRSI	KARNATAKA
EJN0528407	Kamal	JAGJEET	21-Oct-74	MODEL TOWN NARNAUL	HARYANA
WCE1298661	Nisha Dhankher	ISHVAR	11-Jan-59	ANANDUR	TAMIL NADA
OJN0533258	Pankaj	SURENDER SINGH	27-Apr-65	KILLIYUR	KERALA
KMM1008714	Anji Dahiya	MANOJ KUMAR	20-Aug-84	ADONI	ANDHRA PRADESH
MJE1191908	Ani Dahiya	DINKAR KUMAR	01-Feb-76	NEDUMANGAD	KERALA
TPN0860674	Anu Khana	ARUN KHANNA	16-Nov-82	BARMANA	HIMACHAL PRADESH
USP2119177	Muskaan Solanki	DINESH SOLANKI	15-Jul-80	JETPUR	GUJRAT
KCE0252111	Nisha Bansal	SANDEEP BANSAL	20-Oct-65	FAIZPUR NAGPUR	MAHARASHTRA
XHV0405054	Karan	RAVINDER ANTIL	18-Jun-71	VAVDI RAJKOT	GUJRAT
GHY1023521	Veer	JAGAT BIHARI LAL	23-Jul-73	ANANTAPUR	ANDHRA PRADESH
AJU1176677	Anuj	MOHAMMAD ALI	27-Mar-61	ALWAR	RAJASTHAN

Fig4. Database for Prior Approval Check

Figure 4 shows the prior approval check dataset which is used to give information about the banned persons in the country i.e. which are not eligible for applying for new passport. In this a separate list for banned people are considered for this check and score is assigned on the basis of this database.

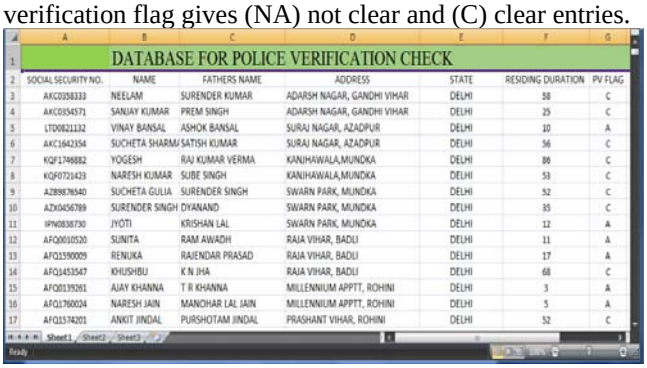
Figure 5 shows the database for the index check which checks the old passport number and its date of issue and expiry. In this dataset an IC flag is raise in the form of NC (not clear) and C (clear). Not clear flag is raise for the entry which is not eligible for issuing a new passport.

DATABASE FOR INDEX CHECK

SOCIAL SECURITY NO.	NAME	FATHER NAME	DATE OF BIRTH	ADDRESS	OLD PASSPORT NO.	IC FLAG
ABC5715726	SAPNA GIRDHAR JHURANI	THANWARDAS JHAMUMAL C	27-Jan-64	CHENNAI	PP10010	NC
PQR1234567	RITA CHOWDHRI	DEVINDER KUMAR JAIN	31-Oct-68	BHIWANI	PP10011	NC
STU8756942	QAMARUDDIN	RIAZUDDIN	10-May-71	SITAPUR	PP10012	C
DEF2356489	BIJENDER CHANDELIYA	RAJ KUMAR	13-Oct-76	SONIPAT	PP10013	NC
TUV0854860	MAYUR WADHWA	SURESH CHANDRA WADHWA	29-2-1979	SECTOR C VASANT KUNJ	PP10014	NC
UVW0212520	GAGANDEEP SINGH	RAMINER SINGH	03-Aug-79	MODEL TOWN 3RD	PP10015	C
XYZ1534009	DUKHTARI BEGUM	NIZAM ALI KHAN	19-Nov-70	PITAMPURA DELHI	PP10016	NC
KLM1450547	SARVESH KUMAR SINGH	RAM PRATAPNSINGH	11-Dec-74	SADH NAGAR PALAM	PP10017	NC
QRS0450261	GOPALAKRISHNA PILLAI	KRISHNA KURUP	10-Mar-59	KURATHIKADU ALLEPPEY		NC
FGH1661024	SEEMA RANI	RAM SNEHI	17-Jun-65	QAZI HOZ	PP10019	NC
YZA1532201	NEELSHA BHERWANI	SHIV KUMAR BHAMBHANI	24-Oct-74	LAJPAT NAGAR 2	PP10020	NC
DEF2108719	SEEMA RANI	DINESH MOHAN BANSAL	13-Nov-72	SECTOR 8 ROHINI	PP10021	NC
JKL1567652	NAVEEN CHAUDHRY	RAM LAL CHAUDHARY	13-Aug-80	MAHAVIR NAGAR TILAK N	PP10022	NC
MNO1571738	SANJIV KUMAR BAHADUR	AJOY KUMAR BAHADUR	08-Dec-64	PATLIPUTRA COLONY PATI	PP10023	C
BCD1664202	AVNISH KUMAR GUPTA	RAM AVTAR GUPTA	27-Apr-72	KHERU MUCHAND	PP10024	NC
HIJ2534215	JAGPAL SINGH	HUKAM SINGH	19-Apr-66	KANOOR GAZIABAD	PP10025	NC

Fig5. Database for Index Check

Figure 6 shows the dataset for Police Verification Check contains basic information about the person which was applied for the passport. This database has attributes like a unique social security number, applicant name, fathers name, date of birth, residing duration and a police verification flag gives (NA) not clear and (C) clear entries.

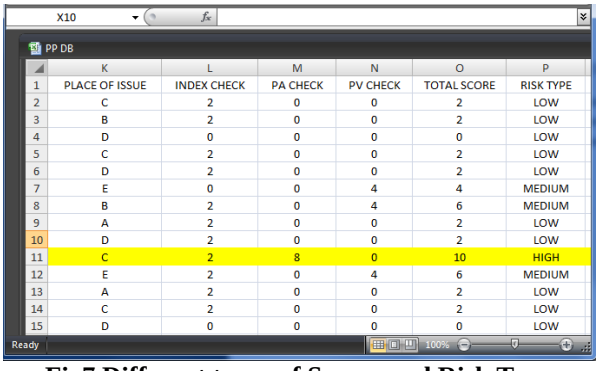


DATABASE FOR POLICE VERIFICATION CHECK

SOCIAL SECURITY NO.	NAME	FATHERS NAME	ADDRESS	STATE	RESIDING DURATION	PV FLAG
AKC0358333	NEELAM	SURENDER KUMAR	ADARSH NAGAR, GANDHI VIHAR	DELHI	58	C
AKC0354571	SANJAY KUMAR	PREM SINGH	ADARSH NAGAR, GANDHI VIHAR	DELHI	25	C
LTD0821132	VINAY BANSAL	ASHOK BANSAL	SURAJ NAGAR, AZADPUR	DELHI	10	A
AKC1642354	SUCHETA SHARMA	SATISH KUMAR	SURAJ NAGAR, AZADPUR	DELHI	56	C
KQF1746882	YOGESH	RAJ KUMAR VERMA	KANJHAWALA,MUNDKA	DELHI	86	C
KQF0721423	NARESH KUMAR	SUBE SINGH	KANJHAWALA,MUNDKA	DELHI	53	C
AZB8876540	SUCHETA GULIA	SURENDER SINGH	SWARN PARK, MUNDKA	DELHI	52	C
AZX0456789	SURENDER SINGH	DYANAND	SWARN PARK, MUNDKA	DELHI	35	C
IPN0838730	JYOTI	KRISHAN LAL	SWARN PARK, MUNDKA	DELHI	12	A
AFQ0010520	SUNITA	RAM AWADH	RAJA VIHAR, BADLI	DELHI	11	A
AFQ1590009	RENUKA	RAJENDAR PRASAD	RAJA VIHAR, BADLI	DELHI	17	A
AFQ1453547	KHUSHBU	K N JHA	RAJA VIHAR, BADLI	DELHI	68	C
AFQ0139261	AJAY KHANNA	T R KHANNA	MILLENNIUM APPTT, ROHINI	DELHI	3	A
AFQ1760024	NARESH JAIN	MANOHAR LAL JAIN	MILLENNIUM APPTT, ROHINI	DELHI	5	A
AFQ1574201	ANKIT JINDAL	PURSHOTAM JINDAL	PRASHANT VIHAR, ROHINI	DELHI	52	C

Fig6. Database for Police Verification

Figure 7 shows the score for different checks like index check, prior approval check and police verification check then with the help of these scores total score was calculated and then classify the risk according to the total score of the entry and the risk type was assigned to each entry in the database.



PP DB

PLACE OF ISSUE	INDEX CHECK	PA CHECK	PV CHECK	TOTAL SCORE	RISK TYPE
C	2	0	0	2	LOW
B	2	0	0	2	LOW
D	0	0	0	0	LOW
C	2	0	0	2	LOW
D	2	0	0	2	LOW
E	0	0	4	4	MEDIUM
B	2	0	4	6	MEDIUM
A	2	0	0	2	LOW
D	2	0	0	2	LOW
C	2	8	0	10	HIGH
E	2	0	4	6	MEDIUM
A	2	0	0	2	LOW
C	2	0	0	2	LOW
D	0	0	0	0	LOW

Fig7.Different types of Scores and Risk Type

6. CONCLUSION

This paper gives the conceptual information regarding to the computation and identification of the risk in the passport data analysis. This work helps in identifying the activities, performed by applicants registered for issuing the passport, which are illegal or fraud. This describes the work flow and the procedure for calculating the risk score for each entry in the database. In this different types of checks like index check, police verification check and prior approval checks are described which are also used by the passport offices for finding the fraud entries in the database. The attributes used in the passport database are the critical attributes which are prone to the risk.

7. FUTURE WORK

This only gives the conceptual information regarding to the computation of score and classifying risk. The computed score and identified risks are also used for distributing the risk to check from where or which office the sensitive data originates. Furthermore the results will also validate using different algorithms.

ACKNOWLEDGEMENT

Authors would like to thanks to their head Dr. Rajan Vohra, Head Of Department of CSE & I.T department, PDMCE, Bahadurgarh, India for his valuable support and help.

REFERENCES

[1] HAND, D. J. (1981). Discrimination and Classification. Wiley, Chichester

[2] MCLACHLAN, G. J. (1992). Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York.

[3] RIPLEY, B. D. (1996). Pattern Recognition and Neural Networks.Cambridge Univ. Press.

[4] HAND,D. J. (1997). Construction and Assessment of Classification Rules. Wiley, Chichester.

[5] WEBB, A. R. (1999). Statistical Pattern Recognition. Arnold, London.

[6] CLARK, P. and NIBLETT, T. (1989). The CN2 induction algorithm. Machine Learning 3 261–285.

[7] QUINLAN, J. R. (1990). Learning logical definitions from relations. Machine Learning5 239–266.

[8] QUINLAN, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

[9] COHEN, W. (1995). Fast effective rule induction.In Proceedings of the 12th International Conference on Machine Learning 115–123.Morgan Kaufmann, Palo Alto, CA.

[10] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). Classification and Regression Trees.Wadsworth, Belmont, CA.

[11] CHAN, P. K., FAN, W., PRODROMIDIS, A. L. and STOLFO, S. J. (1999). Distributed data mining in credit card fraud detection. IEEE Intelligent Systems 14(6) 67–74.

[12] Lefait, G. and Kechadi, T, (2010) "Customer Segmentation Architecture Based on ClusteringTechniques" Digital Society, ICDS'10, Fourth International Conference, 10-02-2010.

[13] Fraley, Andrew, and Thearting, Kurt (1999). Increasing customer value by integrating data mining and campaign management software. DataManagement, 49–53.

[14] P. Bhargavi and S.Jyothi, (2009) "Applying Naïve Bayes Data Mining Technique for Classification of Agricultural land Soils" IJCSNS International Journal of computer Science and Network Security, VOL. 9 No.8, August 117-122.

[15] I.Krishna Murthy, "Data Mining- Statistics Applications: A Key to Managerial Decision Making",Article/Report indiastat.com, April-May 2010.

[16] Association Analysis of Customer Services from the Enterprise Customer Management System- ICDM-2006.

[17] Database collected from National Informatics Centre nodal office for passport preparation in New Delhi, INDIA.