

b'2015 International Conference on Communication, Information & Computing Technology (ICCICT), Jan. 1

b' 978-1-4799-5522-0/15/\$31.00 \xc2\xa92015 IEEE\n'

b' I am testing Vipula Rawte \n'

b'Student, M.E. (Computer Engineering) \n'

b'St.Francis Institute of Technology \n'

b'Mumbai-400103, India. \n'

b'rawtevipula25@gmail.com \n'

b'G Anuradha \n'

b'Associate Professor (Computer Engineering) \n'

b'St.Francis Institute of Technology \n'

b'Mumbai-400103, India. \n'

b'ganusrinu4@yahoo.co.in\n'

b' Abstract\n'

b'\xc5\xa0Fraud is widespread and very costly to the health-\n'

b'care insurance system. Fraud involves intentional deception or \n'

b'misrepresentation intended to result in an unauthorized benefit. \n'

b'\n'

b'It is shocking because the incidence of health insurance fraud \n'

b'keeps increasing every year. In order to detect and avoid the \n'

b'fraud, data mining techniques are applied. This includes some \n'

b'\n'

b'preliminary knowledge of health care system and its fraudulent \n'

b'\n'

b'behaviors, analysis of the characteristics of health care insurance \n'

b'data. Data mining which is divided into two learning techniques \n'

b'viz., supervised and unsupervised is employed to detect \n'

b'\n'

b'fraudulent claims. But, since each of the above techniques has its \n'

b'\n'

b'own set of advantages and disadvantages, by combining the \n'

b'advantages of both the techniques, a novel hybrid approach for \n'

b'detecting fraudulent claims in health insurance industry is \n'

b'\n'

b'proposed. \n'

b'Keywords\xc5\xa0data mining; health insurance fraud; \n'

b'supervised; unsupervised \n'

b'I. INTRODUCTION \n'

b' Deliberately deceiving the health insurance company that \n'

b'results in healthcare benefits being paid illegitimately to an \n'

b'\n'

b'individual or group is known as health insurance fraud. The \n'

b'\n'

b'main purpose of fraud is financial benefit. According to a \n'

b'\n'

b'recent survey, it is estimated that the number of false claims in \n'

b'\n'

b'the industry is approximately 15 per cent of total claims. \n'

b'\n'

b'Insurance companies in USA incur losses over 30 billion USD \n'

b'\n'

b'annually to healthcare insurance frauds. The statistics is \n'

b'\n'

b'appalling in developing country like India as well. The report \n'

b'\n'

b'suggests that the healthcare industry in India is losing \n'

b'\n'

b'approximately Rs 600-Rs 800 crores incurred on fraudulent \n'

b'\n'

b'claims annually [1]. Frauds blow a hole in the insurance \n'

b'\n'

b'industry. Health insurance is a bleeding sector with very high \n'

b'claims ratio. So, to make health insurance industry free from \n'

b'fraud, it is necessary to focus on elimination or minimization \n'

b'\n'

b'of fake claims arriving through health insurance. \n'

b'The health insurance fraud claims are broadly classified \n'

b'under the following headings: \n'

b'\xe2\x80\xa2 Billing for services not rendered: Billing insurance \n'

b'company for things that never happened. Example: \n'

b'\n'

b'Forging the signature of those involved in giving bills. \n'

b'\xe2\x80\xa2 Upcoding of services: Billing insurance company for \n'

b'services that are costlier than the actual procedure that \n'

b'\n'

b'was done. Example: 45-minute session being billed as \n'

b'\n'

b'60-minute session \n'

b'\xe2\x80\xa2 Upcoding of items: Billing insurance company for \n'

b'medical equipment that is costlier than the actual \n'

b'\n'

b'equipment. Example: Billing for power assisted \n'

b'\n'

b'wheelchair while giving the patient only the manual \n'

b'\n'

b'wheelchair. \n'

b'\xe2\x80\xa2 Duplicate claims: Not submitting exactly the same bill, \n'

b'but changing some small portion like the date in order \n'

b'\n'

b'to charge insurance company twice for the same \n'

b'\n'

b'service rendered. Example: An exact copy of the \n'

b'\n'

b'original claim is not filed for the second time, but \n'

b'\n'

b'rather some portion like date is changed to get the \n'

b'\n'

b'benefit twice the original. \n'

b'\xe2\x80\xa2 Unnecessary services: Filing claims which in no way \n'

b'apply to the condition of a patient. Example: Patient \n'

b'\n'

b'with no symptoms of diabetes filing claim for daily \n'

b'\n'

b'usage of insulin injections. \n'

b'II. DATA \n'

b'MINING\n'

b' Nowadays there is huge amount of data stored in real-\n'

b'world databases and this amount continues to grow fast. So, \n'

b'there is a need for semi-automatic methods that discover the \n'

b'hidden knowledge in such database. Data mining \n'

b'\n'

b'automatically filtering through immense amounts of data to \n'

b'\n'

b'find known/unknown patterns, bring out valuable new \n'

b'\n'

b'perceptions and make predictions. \n'

b'Data mining techniques tend to learn models from data. \n'

b'There are two approaches on learning the data mining models. \n'

b'\n'

b'Those are supervised learning, unsupervised learning; and \n'

b'\n'

b'they are described below: \n'

b'A. Supervised Learning: \n'

b'This is the most usual learning technique wherein the \n'

b'model is trained using pre-defined class labels. In the context \n'

b'\n'

b'of health insurance fraud detection the class labels may be the \n'

b'\n'

b'\xef\xac\x81legitimate\xef\xac\x82 and \xef\xac\x81fraudulent\xef\xac\x82 claims. The training dataset can