

# **ONLINE PAYMENTS FRAUD** **DETECTION USING WITH MACHINE** **LEARNING:**

**To build an application that can detect the legitimacy of the transaction in real-time and increase the security to prevent fraud.**

By

***(Marri yashmitha)***

***(Manthina raja rishika)***

***(Kutagulla safa)***

*Guided by*

***Prof. Ms swetha raj***

A Dissertation Submitted to  
SRI VENKATESWARA COLLEGE OF  
ENGINEERING AND TECHNOLOGY, An  
Autonomous Institution affiliated to  
‘JNTU Ananthapur’ in Partial Fulfilment of  
the Bachelor of Technology branch of  
***Computer science and Engineering***

*May 2024*



# **SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY**

**R.V.S. Nagar Tirupathi Road, Andhra Pradesh– 517127**

## **Feature selection for online fraud detection:**

Feature selection is an important aspect of machine learning as it greatly affects how the machine learning model performs. Here critical features from a dataset are selected to train the classifier model. This has been seen to improve the performance of the machine learning model. Other benefits of this step include reducing overfitting which means reducing the noise in a dataset. It also reduces training time. As such, it is important to discover the relevant features in a dataset using feature selection. This will help to provide better outcomes of the classifier model. In developing credit card fraud detection models, feature selection techniques are used to find out the most relevant features that best distinguish between legitimate and fraudulent transactions. Several techniques exist for feature selection, some examples are correlation-based feature selection, recursive feature elimination, and univariate selection. In recursive feature elimination, each column is measured based on its importance in relation to the labelled column. Here, the columns with the least importance are eliminated. The process is repeated until only those columns which are important are identified. (Yan and Zhang, 2015) identified that this feature selection technique improved the effectiveness of a binary

classifier when trained with synthetic gas sensor data. However, it was found to have some biases when used in the SVM classifier.

Additionally, (Emura, Matsui, and Chen, 2019) used univariate selection. Here, statistical tests are applied to identify the top features in the dataset. Each column is then assigned a score according to how they performed in identifying the labelled column. (Karegowda, Manjunath and Jayaram, 2010) applied the correlation-based feature selection technique in their experiment. Here, the correlation between each column and the labelled column is measured. A correlation matrix is produced which shows and ranks the relationship between all features. With this, the features with correlation to the labelled column can be deduced as well as how high or low the correlation is. (Kumar, et al., 2019) also used confusion matrix and random forest in developing a credit card fraud classifier model. After evaluating the model, they reported that the performance of the model using the correlation-based feature selection achieved better results of about 90% accuracy. This paper also used this feature selection technique to identify features that are important in modelling. The next section highlights this paper's research methodology.