IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 4, December 2023, pp. 1985~1994
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i4.pp1985-1994

Utilizing deep learning, feature ranking, and selection strategies to classify diverse information technology ticketing data effectively

Mudragada Venkata Subbarao[1], Kasukurthi Venkatarao[1], Suresh Chittineni[2], Subhadra Kompella[2]
[1]Department of CS & SE, AUCE(A), Andhra University, Visakhapatnam, India
[2]Department of Computer Science & Engineering, Gitam University, Visakhapatnam, India

Article Info

Article history:

Received Oct 10, 2022
Revised Feb 26, 2023
Accepted Mar 10, 2023

Keywords:

Deep learning
Incident response
Machine learning
Support vector machine
Text mining

ABSTRACT

In today's internet world, information technology (IT) ticketing services are potentially increasing across many corporations. Therefore, the automatic classification of IT tickets becomes a significant challenge. Feature selection becomes most important, particularly in data sets with several variables and features. However, enhance classification's precision and performance by stopping insignificant variables. Through our earlier research, we have categorized the unsupervised ticket dataset. As a result, we have converted the dataset into a supervised dataset. In this article, the classification of different IT tickets on Machine learning algorithms, Feature ranking, and feature selection techniques are used to improve the efficiency of machine learning algorithms. However, compared to the machine learning (ML) algorithms, the convolutional neural network (CNN) algorithm provides a better classification of the token IDs and provide better accuracy.

Corresponding Author:

M. Venkata Subbarao
Department of CS and SE, AUCE(A), Andhra University
Visakhapatnam, India
Email: subbaraomudragada@gmail.com

1. INTRODUCTION

While the global economy has focused on services rather than products, technological advancements have kept place. Because of the wide range of electronic platforms that offer services, Information Technology IT has become a vital part of our daily lives [1]. Many people use them for leisure, shopping, and other activities. Every company now has a collection of applications that have evolved due to digitization. A large and complicated IT infrastructure is needed to support this product line [2], These advances demonstrate that IT support systems are critical in an organization's support operations. In contrast, huge organizations spend millions of dollars on commercial text classification algorithms for small enterprises. IT company workers face various difficulties, including challenges with buildings and infrastructure, software, and HR issues. The IT service desk or Helpdesk, which is often accessible over the Internet, is used by employees of an organization to report an issue [3]. The problem tickets will be assigned to the relevant domain expert group or service desk representative based on the ticket category.

Ticket categories, priority, and severity are just a few of the structured fields in web-based IT service desk solutions [4]. A free-form field called "ticket description" allows the user to submit a description of the ticket in their language. Employees manually select the problem's category, priority, and severity, as well as its description in standard English, while creating trouble tickets. Manual selection of the ticket category by the end user may lead to an incorrect ticket classification because it is based on the user's impression of the problem and if the user has registered the issue in the relevant category [5]. When tickets are incorrectly categorized, they will be sent to the wrong resolution group, which will cause a delay in resolving the issue tickets. Conventional service desk systems work best with well-structured datasets [6]. We can use a variety of machine-learning approaches to build an automatic ticket classification system that addresses all of these issues. For example, to categorize a service desk ticket, an automated ticket classifier analyses the ticket's end-description user in natural language, which uses both supervised and unsupervised machine learning approaches to build ticket classifier models [7]. Furthermore, classifier models can be constructed using supervised machine learning techniques such as classification algorithms when the label or category of historical ticket data is known [8]. Therefore, this paper proposes a machine-learning-based classification of IT tickets.

2. RELATED WORK

The information will most likely be presented as free text if a ticket is generated automatically. Harun *et al.* mentioned that the strategy for constructing an automated service desk ticket classifier system was created by Paramesh *et al.* [9], who did their research by analyzing data from IT infrastructure helpdesks. Traditional supervised machine-learning methods are used to construct classification models [10]. A comprehensive investigation is carried out into the methods that can be used to deal with undesirable, imbalanced, or wrongly labeled data. The convolutional neural network performed significantly better than any other model examined in other classification models [11]. Machine learning and natural language processing techniques are utilized during a system's development [12]. Analyze the tickets by manually evaluating them using n-gram analysis and contextual mining [13]. The resulting model had an error rate of only 1.4% when classifying each ticket into the appropriate root cause category. Developed trouble miner to sort trouble tickets according to their underlying causes [14]. According to the results of the study, most tickets are caused by disruptions in the network cables and routers. Constructed regression and classification models to predict the resolution times [15]. It was decided to eliminate the fields that held text data because the text had to be entered by a human every time. There is also a text area included in this thesis; however, the text within it is not produced by a person but rather by a machine. Because of this, the text box would be mined for helpful information. Regarding classification, the resolution time was divided into three categories, and the resulting model had an accuracy of approximately 74.5%. On the other hand, when it came to regression, the artificial neural network had the lowest mean absolute error, which was 24.8 hours [16]. Sample *et al.* mention that according to Lofgren [17], employed data mining and machine learning methods to determine the underlying cause of network issues. This allowed him to provide engineers with actionable recommendations and, as a result, reduce the amount of time spent on the process of troubleshooting. The model had an accuracy of up to 90 percent when predicting the root cause of the most prevalent root cause and only 70% when discriminating between up to 20 different root causes [18].

3. METHOD

Applying the predictive models, feature ranking, and selection techniques to the dataset with the body attribute and also without the body attribute [19]. Preprocessing was carried out to convert the textual data in the body tag to numerical data. To perform this, we have to use the count vectorizer library [20]. After the conversion is done, normalize all the data into a range of 0 to 1. After this step, feature ranking is carried out to understand which features are of utmost importance, and the feature selection technique is used to improve the efficiency of the predictive models such as the support vector machine classifier (SVM/SVC), Gaussian Naive Bayesian, decision trees, logistical regression, and k-nearest neighbours (KNN). The settlement curves produced at SG1 and SG2 has been illustrated in Figure 1.

3.1. Dataset

This dataset was retrieved after performing clustering and labeling mechanisms obtained from our previous study. The best resultant algorithm of the prior survey, latent dirichlet allocation-based (LDA-based) topic prediction, which contains the 13 topics, was used as the target attribute for classification. The dataset includes a total of 47,664 incidents initially taken from the service now platform [21]. Figure 1 and Table 1 show how the characteristics in the dataset were used to perform this study. Yes/No values in the usage column suggest the usage of a particular feature for this study.

3.2. Environment

To conduct this research, we have used the following experimental setup. We have used Python programming language and Jupiter notebook. Also, an Intel i5 processor with 32 GB RAM was used to conduct this study.

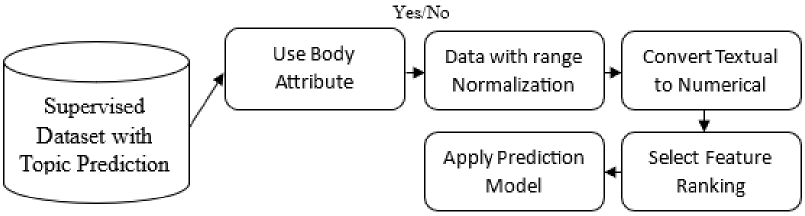

Figure 1. The flowchart of the proposed approach

4. METHODOLOGY

4.1. Attributes in the data set

The body attribute contains textual data, whereas all other fields contain numerical data. Hence, a separate analysis was conducted to identify the performance of classification algorithms using the body attribute [22]. The range for logistic regression is between 0 and 1, but the range for linear regression is unbounded. This is the primary distinction between the two types of regression [23]. In addition, in contrast to linear regression, logistic regression does not mandate the existence of a linear connection between the variables that serve as inputs and those that are analyzed as outputs.

Table 1. Attributes in the data set

Attributes	Description	Usage
Topic prediction	This contains the topic prediction values ranging from 1 to 13	Yes
Body	This field contains the agent entry of the ticket description	*
Ticket type	This field contains a Numerical Value of either 0 or 1, 0 refers to email, and 1 relates to phone	Yes
Category	This field contains a numerical value ranging from 0 to 12	Yes
Sub_category1	This field contains a numerical value ranging from 0 to 58	Yes
Sub_category2	This field contains a numerical value ranging from 0 to 118	Yes
Business service	This field contains a numerical value ranging from 0 to 102	Yes
Urgency	This field contains a numerical value ranging from 0 to 3. 3 is the urgent ticket, 0 no urgency.	Yes
Impact	This field contains a numerical value ranging from 0 to 4. 5 is the highest impact and 0 is the lowest.	Yes

4.2. CNN

It is one of the deep learning algorithms that takes input data and assigns them tags and ids based on their weights or parameters. These tags and ids are used to differentiate the characteristics of the features extracted from the data. It also requires very less pre-processing of the data, as it classifies them by itself during the process and learns from them [24]. The functioning of the convolutional neural network (CNN) algorithm as shown in Figure 2 is similar to that of the human brain [25]. It consists of neurons that pass through several networks to modify the extracted data and finally learn the features [26]. It also takes advantage of the spatial and temporal features of the dataset and improves its functioning. Mostly, it is suitable for image datasets as it takes advantage of the pixel information in the images. The layers used in the CNN algorithm are discussed in detail in the following table.

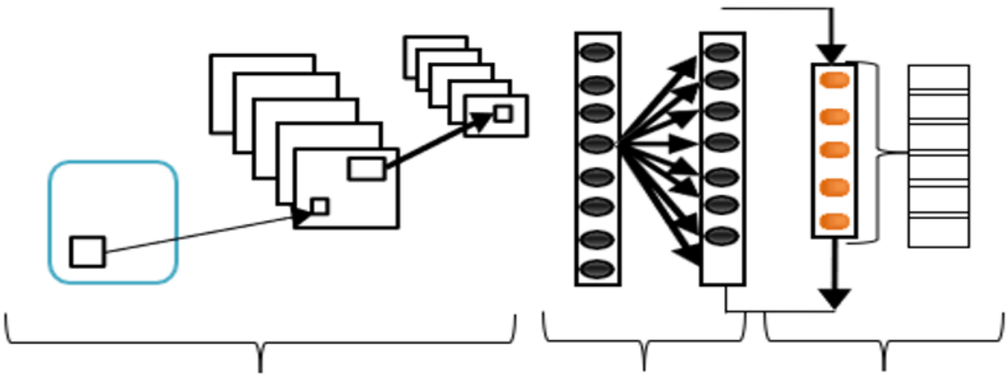

Figure-2. The architecture of the CNN algorithm

Let CNN have in m blocks Let us, input x represented as the probability of getting connected to the ith block of the SAM [27]. A controller output helps in sampling the realization a which is a~p. The scheme a has the probability of p^ = (p^1, p^2, …. , p^m). The G(a) is denoted as input of the lth block and the mapping of the residual is $fl(.\ )$, then the xl+1 is the output of the lth block is derived,

$$x_{l+1} = x_l + f_l(x_l) \tag{1}$$

– **FULL SELF-ATTENTION (FULL-SA) NETWORK**

In (2) shows that l^{th} block is denoted as $M(.\ ; wl)$, and it is placed in wl parameters. Then the equation is developed $M(fl(xl);\ Wl)$ which includes the processing end of the extraction process [28]. Finally, the residual output is $fl\ (xl)$. Where l = 1, …., m and ⊙ is denoted the element-wise multiplication. In (2) defines the cost of the computation and the parameters are increased based on the number of blocks m.

$$x_{l+1} = x_l + M(f_l(x_l) \colon W_l) \odot f_l(x_l) \tag{2}$$

– **CONNECTION SCHEME**

Assume that the CNN has m blocks. A sequence $a = (a1,\ a2\ \ldots,\ am)$ indicates a connection scheme, where ai = 1 and i^{th} block is connected to a SAM. The scheme formulated in (3) is given.

$$x_{l+1} = x_l + (a_l + M(f_l(x_l) \colon W_l) + (1 - a_l).1) \odot f_l(x_l) \tag{3}$$

Al the one vector is defined here as 1 and the l lies between the 1 and m. The all-one vector is represented as a full-SA network, and the CNN allows the neurons if a represents 0. The reward is given to a. The controller has the parameter set of θ and the η is the policy gradient of the learning rate.

$$R_\theta = G(a).\sum_{t=1}^{m} logp_\theta^{\wedge i}, \theta \leftarrow \theta + \eta.\nabla R_\theta \tag{4}$$

In this manner, the controller provides the probability for the reward G. Searching for a good structure of G can help in finding a better structure [28]. Through the connection ratio and accuracy, the better reward G can be found. The subnetwork $(x \vee a)$ provides a validation accuracy of $gval$ which is obtained by sampling the super net of the reward.

$$\frac{I_t\,(CNN\ with\ SAMs) - I_t\,(Original\ CNN)}{I_t(Original\ CNN)} \times\ 100\% \tag{5}$$

The network's inference time It (.). The batch size is defined to be between 50 to 1,000 times. The $T(x)$ is defined as the Lipschitz continuous function, and the K is the d-dimensional compact set's Lebesgue integrable function [29]. The overall subnetwork consists of the depth and width of the layer $Rful(x,\ \theta full)$ which is smaller than the constant ϵ0.

$$\int_K |R_{full}(x, \theta_{full}^0)\ - R_{sub}(x)|\, dx \leq \epsilon \tag{6}$$

$$\int_{R^d} |f(x)\ - R(x)|\, dx \leq \epsilon \tag{7}$$

$$\int_{R^d} |f(x, \theta_f^0)\ - T|\, dx \leq \epsilon \tag{8}$$

The skip connections are seen from the formulas, $(x,\ \theta 0) = f(x,\ \theta 0)$.

$$\int_{R^d} |f(x, \theta_f^0)\, T|\, dx = \int_{R^d} |g(x, \theta_g^0)\ - T|\, dx \leq \epsilon \tag{9}$$

$$\omega_K(r)\ = \max_{x,y\, \epsilon\, K, ||x-y|| \leq\, r} |f(x)\ - f(y)| \tag{10}$$

Let T(x) can be seen that Lipschitz continuous function algorithm, in which the following equation shows its functioning,

$$|T(x) - T(y)| \leq L|x - y| \tag{11}$$

then we have given that $wk(r), = Lr = \epsilon/Vol\ (K)$.

$$r = \frac{\epsilon}{Vol(K)\cdot L} \tag{12}$$

Here, if the r= %epsilon/ (Vol(k).L) , then the $\epsilon \equiv (\epsilon_{0,1})$,

$$O(1/r^d) = O\left(\left(\frac{L}{\epsilon}\right)^d\right) < O\left(\left(\frac{L}{\epsilon_0}\right)^d\right) = C\left(\frac{L}{\epsilon_0}\right)^d \tag{13}$$

the constant C in the equation is multiplied by the Lemmavalue, which exists in the CNN $R_{short}(x)$ and its width lies within d and the $C(L/\epsilon_0)^d$ also lies within it.

$$dep(R_{long}) = dep(R_{full}) \tag{14}$$

It can be seen from the equation that; the d value is greaterthat $R_{lon.}$ It can also be stated that $x\epsilon K$, and $Rshort(x)$.

$$R_{long}(x) = \int_k \left|T(x) - R_{long}(x)\right|dx = \int_k \left|T(x) - R_{short}(x)\right|dx \leq \epsilon/2 \tag{15}$$

$$Then \int_k \left|R_{full}(x, \theta_{full}^0) - R_{long}(x)\right|d \leq \int_k \left|T(x) - R_{long}(x)\right|dx \tag{16}$$

$$+ \int_k \left|R_{full}(x, \theta_{full}^0) - T(x)\right|dx \tag{17}$$

$$\leq \epsilon/2 + \epsilon_0/2 \leq \epsilon \tag{18}$$

The $Rlong$ represents the CNN algorithm that has a width smaller than that of d, while the $Rfull$ is greater than d. The inequality in the network is satisfied by the $Rfull$.

$$m_c^l = AVG(X_c^l) = \frac{1}{H\cdot W}\sum_{h=1}^{H}\sum_{w=1}^{W} X_{chw}^l \tag{19}$$

The features are processed by the sigmoid function si(z) = 1⁄(1 + e−z). Here the reduction rate is defined as r and the division extracted as '//'. Thus, the size of the hidden layer is C//r. The information obtained from all the channels is fused using the rectified linear unit (ReLU) activation function [30]. The block wise information of the long short-term memory (LSTM) is integrated through the environmental impact assessment (EIA) module. The average pooling output is termed as ml that is passed to the hidden state h.

$$[\delta_1; \cdots; \delta_C] = sig(FC([m_1^l; \cdots; m_C^l]; W_l)) \tag{20}$$

The block wise information of the LSTM is integrated through the EIA module. The average pooling output is termed as ml that is passed to the hidden state h. Zero vectors are termed as the h0 and c0.

$$(h_l, c_l) = LSTM([m_1^l; \cdots; m_C^l], h_{l-1}, c_{l-1}; W) \tag{21}$$

The G represents the number of groups, and the feature maps are represented as C/G [31]. The feature maps are grouped as $(C//G) \times H \times W$ that represents the Yl that lies within X^l.

$$g_c^l = AVG(Y_c^l) = \frac{1}{H\cdot W}\sum_{h=1}^{H}\sum_{w=1}^{W} Y_{chw}^l \tag{22}$$

The coefficient of importance for each value g in [g ; …; gl] 1 $C//G$,

$$p_{hw} = g.Y[:, h, w] \tag{23}$$

the value p_hw is normalized in the following steps.

$$\hat{p}_{hw} = \frac{p_{hw} - \mu}{\sigma + \epsilon} \tag{24}$$

The μ and $\sigma 2$ are the mean and variance of the tickets which can be calculated through the following equation,

$$\mu = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} p_{hw}\, , \sigma^2 = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(p_{hw} - \mu^2) \tag{25}$$

for the group Yl, some additional set of parameters (γ, β) are added to rescale and normalize the features, and the attention received by the sun grid engine (SGE) modules [:, h, w] are written as,

$$sig(\gamma p_{hw} + \beta) \tag{26}$$

the full CNN algorithm is accelerated through the g_spa. The g_spa encourages the network to achieve better results by generating fewer connection schemes that are found between the tickets.

$$g_{spa} = 1 - \frac{\|a\|_0}{m} \tag{27}$$

More schemes can be explored in this process, as there is mitigation and convergence during the training iterations and the reassigned number database (RND) bonus provides better convergence of the iterations. The difference in the output is reduced by the RND.

$$G(a) = \lambda_1 . g_{spa} + \lambda_2 . g_{val} + \lambda_3 . g_{rnd} \tag{28}$$

The proximal policy optimization method is used for faster training and sampling of the connection schemes. It also provides better efficiency in the utilization of the data [27]. The tuple is kept in a buffer after updating the parameters %theta and %phi. Layer(type):conv2d_1, conv2d_2 max_pooling2d_1, dropout_1, dropout_3, dense_2 Output:(Conv2D),(Conv2D), MaxPooling2,(Dropout,(Dropout), (Dense) Shape:(None, 75, 100, 32),(None, 75, 100, 32), (None, 37, 50, 32,) (None, 37, 50, 32), (None, 128), Parameters: 896, 9248, 0, 0, 903. Total params: 3,752,999, Trainable params: 3,752,999, non-trainable params: 0

The number of convolutional, pooling, and fully convoluted layers used in the algorithm [32]. It also is said that the proposed algorithm provides better ticketing of the IT and it needs to be confirmed whether the proposed algorithm is more efficient than other existing methods. Most of the existing methods considered classification algorithms, while this paper has considered CNN for the efficient processing of the data.

4.3. Feature ranking and selection

Feature selection as automatically contributes the most prediction variable or output in which you are interested which is made by feature extraction [33]. The following are some of the advantages that come from completing feature selection before modeling your data: To avoid overfitting, collect fewer duplicate data. This will offer the model a performance boost and result in fewer opportunities to make judgments based on noise. In addition, it reduces the amount of time needed for training. Since there is fewer data, the algorithms train more quickly [34].

4.4. Chi-square

In statistics, the chi-square test is used to determine whether or not two occurrences may be considered independent. Chi-square, we utilize it in the feature selection process to determine whether or not the incidence of a specific word and the occurrence of a particular class are independent of one another [35]. Oi = Actual Observation Ei = Expectation. If the matching chi-square score for each feature is high, this suggests that the null hypothesis H0 of independence should be rejected and that the occurrence of the feature and class depend on one another [36].

$$\chi 2 = \sum(Oi - Ei)2/Ei \tag{29}$$

4.5. Recursive feature elimination (RFE)

Recursive feature elimination for selecting features that best fit a model and eliminating the part (or features) that are the weakest until the necessary number of features has been attained. The features are prioritized according to the model's coefficient or the feature priority characteristics. RFE makes an effort to remove any dependencies and collinearity present in the model by iteratively deleting a small number of features at each iteration of the loop [37]. RFE necessitates retaining a certain number of features; however, it is not always possible to predict how many elements will be considered legitimate. Therefore, cross-validation is used with RFE to score several feature subsets and choose the collection of features with the highest score. This allows for the optimum features to be determined.

5. RESULT

The category attribute in the dataset consists of 13 categories. All the tickets in the dataset are labeled with topic prediction results from our earlier research work. Tables 2 and 3 show the results of the performance analysis. Table 2 shows the result of recursive feature elimination on the obtained dataset [38]. Logistic regression ranked 1 for urgency, impact, and ticket type attributes [39]. On the other hand, predictive algorithms such as SVC, Gaussian Naïve Bayesian, and KNN algorithms were not applicable with RFE and hence denoted as NA.

Table 2. Features rankings with and without body attribute using RFE

	FEATURE RANKING USING R.F.E. WITH BODY				FEATURE RANKING USING R.F.E WITH BODY			
	Logistic Regression	Random Forest	Decision Trees	CNN	Logistic Regression	Random Forest	Decision Trees	CNN
Ticket Type	1	5	5	6	2	8	8	7
Category	2	3	3	5	5	1	1	1
Subcategory 1	3	1	1	1	6	1	1	1
Subcategory 2	5	1	1	1	8	1	1	1
Business	4	1	1	1	7	1	1	1
Urgency	1	2	2	3	4	1	1	1
Impact	1	4	4	4	3	7	7	6
Body	1	2	2	3	1	(2-6)	(2-6)	(1-5)

The results of using RFE with the body attribute are presented in Table 2. When the boy attribute is used, we can observe a change in the ranks of features in the dataset. Body attribute has 10 details when they are converted from textual to numerical value [40]–[42]. The random forest and decision trees have been awarded a ranking of 2 to 6 for all 10 body attributes. Random forests and decision trees had a similar order for the other features. Logistic regression has awarded a rank1 to the body attribute.

Decision trees had a higher accuracy and better specificity and sensitivity when compared with the logistic regression and random forest while using the body attribute without the body attribute [43]–[45]. We have carried out the chi-square feature selection technique on our obtained dataset. The number of features value is set to 3. Feature selection has produced a similar result as the Feature ranking, where features sub category 1, subcategory 2, and business had better results when compared with any other combination of features [46]–[48]. Table 3 and Figure 2 show the performance analysis results of employing the chi-square feature selection technique on the predictive models. KNN has a better accuracy of 89.22% over the other models when body attribute is not used, and SVM/SVC had a better accuracy of 86.86% compared with the others. Table 3 shows the mean performance of these models. To evaluate the overall better predictive model, we have conducted a mean performance analysis where the average is calculated considering accuracy, specificity, and sensitivity for both with and without body attributes [49], [50].

Table 3. Performance of chi-square on with and without body attribute

Without Body				Without Body			
Methods	Accuracy	Specificity	Sensitivity	Methods	Accuracy	Specificity	Sensitivity
Logistic Regression	87.45	95.65	98.51	Logistic Regression	82.52	92.74	96.31
SVC	87.36	95.44	98.50	SVC	86.86	94.91	97.1
Random Forest	85.26	94.43	97.21	Random Forest	80.81	90.79	94.65
Decision Trees	85.41	94.88	97.56	Decision Trees	81.61	91.56	95.88
Gaussian	87.03	95.03	98.04	Gaussian	82.03	92.15	96.23
KNN	89.22	96.55	98.91	KNN	83.55	93.27	96.59
CNN	98.43	97.56	98.56	CNN	98.32	98.56	98.45

CNN algorithm outperforms all the algorithms computing faster with better accuracy and F1 score. All the algorithms are trained with the training dataset to improve the performance of the algorithms. However, the clustering algorithms provide very less monthly recurring revenue (MRR) compared to the CNN algorithm. It also is termed as the inefficiency of the algorithm to learn the features and it takes time and quality data to improve the accuracy of the algorithms further. But the CNN algorithm with the minimum number of datasets and features provides better classification and MRR. The sample rate tuning of the algorithms is considered in this work. It can be seen Figure 3 that the CNN algorithm has the minimum sample rate tuning compared to other clustering and regressive algorithms. Through this, it can be concluded that the CNN algorithm provides efficient and accurate results compared to other algorithms and it also outperforms other algorithms in terms of cost, resources, and other metrics.

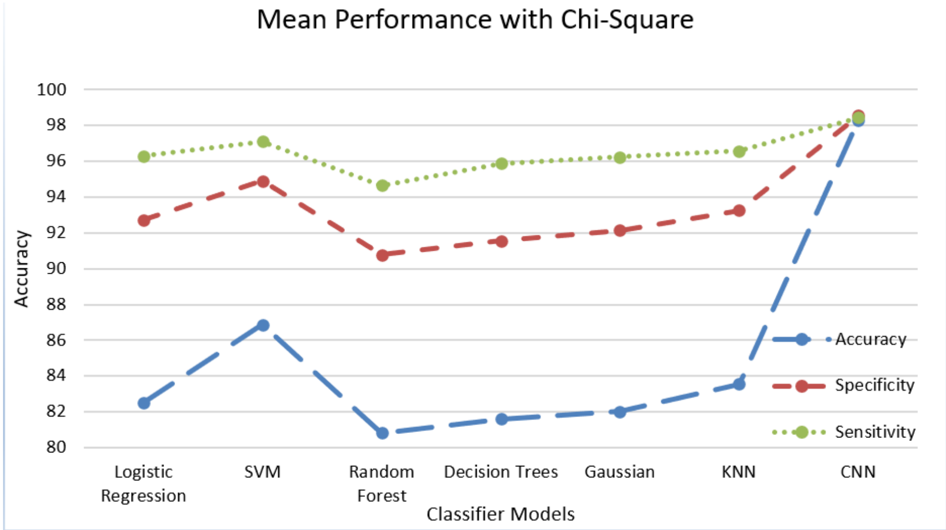

Figure 3. Mean performance of predictive models with chi-square

6. CONCLUSION

As a result of our past work, the unsupervised ticket dataset has been classified and labeled, transforming it into a supervised dataset. In the retrieved dataset, only the body attribute is textual. Through this research, we have conducted performance analysis of several feature ranking and feature selection techniques (RFE and chi-square) when combined with predictive models such as SVM/SVC, Gaussian Naive Bayesian, decision trees, logistical regression, and KNN For Feature ranking, the Decision tree algorithm performed better when compared with the Random Forest or Logistic Regression algorithms. KNN algorithm performed well without using textual data when combined with chi-square. While analyzing the overall performance of predictive models (with and without body attributes), when paired with the chi-square feature selection technique, the CNN algorithm outperformed all other methods with a mean accuracy of 98.32%.

REFERENCES

[1] S. P. Paramesh and K. S. Shreedhara, “Automated IT service desk systems using machine learning techniques,” in *Lecture Notes in Networks and Systems*, vol. 43, Springer Singapore, 2019, pp. 331–346.

[2] S. Agarwal, V. Aggarwal, A. R. Akula, G. B. Dasgupta, and G. Sridhara, “Automatic problem extraction and analysis from unstructured text in IT tickets,” *IBM Journal of Research and Development*, vol. 61, no. 1, pp. 41–52, 2017, doi: 10.1147/JRD.2016.2629318.

[3] J. Xu, H. Zhang, W. Zhou, R. He, and T. Li, “Signature based trouble ticket classification,” *Future Generation Computer Systems*, vol. 78, pp. 41–58, 2018, doi: 10.1016/j.future.2017.07.054.

[4] D. Wang, Y. Liu, X. Qiu, R. Gong, Y. Zhou, and L. Yu, “Anticipated mode generating for on-line verification of relay protection setting based on maintenance tickets information,” *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, vol. 38, no. 19, pp. 81–84, 2014, doi: 10.7500/AEPS20131209014.

[5] G. Son, V. Hazlewood, and G. D. Peterson, “On automating XSEDE user ticket classification,” *ACM International Conference Proceeding Series*, 2014, doi: 10.1145/2616498.2616549.

[6] S. Roy, D. P. Muni, J. J. Y. Tack Yan, N. Budhiraja, and F. Ceiler, “Clustering and labeling IT maintenance tickets,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9936 LNCS, pp. 829–845, 2016, doi: 10.1007/978-3-319-46295-0_58.

[7] S. S. Ali Zaidi, M. M. Fraz, M. Shahzad, and S. Khan, “A multiapproach generalized framework for automated solution suggestion of support tickets,” *International Journal of Intelligent Systems*, vol. 37, no. 6, pp. 3654–3681, Sep. 2022, doi: 10.1002/int.22701.

[8] K. D. Aglibar, G. C. Alegre, G. Del Mundo, K. F. Duro, and N. Rodelas, “Ticketing system: A descriptive research on the use of ticketing system for project management and issue tracking in IT companies,” 2022, doi: 10.25147/ijcsr.2017.001.1.90.

[9] N. A. Harun, S. H. Huspi, and N. A. Iahad, “Question classification framework for helpdesk ticketing support system using machine learning,” Oct. 2021, doi: 10.1109/ICRIIS53035.2021.9617077.

[10] S. Fuchs, C. Drieschner, and H. Wittges, “Improving support ticket systems using machine learning: A literature review,” 2022, doi: 10.24251/hicss.2022.238.

[11] A. Fazari, M. Arnone, C. Botta, B. Caroleo, and S. Pensa, "Mobisuite: A user-friendly tool to exploit e-ticketing data and support public transport planning," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 426 LNICST, pp. 149–161, 2022, doi: 10.1007/978-3-030-97603-3_11.

[12] Y. Temprado, F. J. Molinero, C. García, and J. Gómez, "Knowledge discovery from trouble ticketing reports in a large telecommunication company," *2008 International Conference on Computational Intelligence for Modelling Control and Automation, CIMCA 2008*, pp. 37–42, 2008, doi: 10.1109/CIMCA.2008.116.

[13] N. Deshai and B. B. Rao, "A detection of unfairness online reviews using deep learning," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, pp. 4738–4779, 2022.

[14] N. Deshai, B. V. D. S. Sekhar, P. V. G. D. Prasad Reddy, and V. V. S. S. S. Chakravarthy, "Processing real world datasets using big data hadoop tools," *Journal of Scientific and Industrial Research*, vol. 79, no. 7, pp. 631–635, 2020, doi: 10.56042/jsir.v79i7.40481.

[15] M. V. Subbarao, K. Venkatarao, and C. H. Suresh, "Automation of incident response and it ticket management By Ml and Nlp mechanisms," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 12, pp. 3945–3955, 2022.

[16] A. Medem, M. I. Akodjenou, and R. Teixeira, "Trouble miner: Mining network trouble tickets," *2009 IFIP/IEEE International Symposium on Integrated Network Management-Workshops, IM 2009*, pp. 113–119, 2009, doi: 10.1109/INMW.2009.5195946.

[17] K. R. Sample, A. C. Lin, B. J. Borghetti, and G. L. Peterson, "Predicting trouble ticket resolution," *Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018*, pp. 201–204, 2018.

[18] N. Deshai and B. B. Rao, "Deep learning hybrid approaches to detect fake reviews and ratings," *Journal of Scientific and Industrial Research*, vol. 82, no. 1, pp. 120–127, 2023, doi: 10.56042/jsir.v82i1.69937.

[19] A. Revina, K. Buza, and V. G. Meister, "IT ticket classification: The simpler, the better," *IEEE Access*, vol. 8, pp. 193380–193395, 2020, doi: 10.1109/ACCESS.2020.3032840.

[20] T. R. Eikebrokk and J. Iden, "Strategising IT service management through ITIL implementation: Model and empirical test," *Total Quality Management and Business Excellence*, vol. 28, no. 3–4, pp. 238–265, Aug. 2017, doi: 10.1080/14783363.2015.1075872.

[21] C. Zeng, W. Zhou, T. Li, L. Shwartz, and G. Y. Grabarnik, "Knowledge guided hierarchical multi-label classification over ticket data," *IEEE Transactions on Network and Service Management*, vol. 14, no. 2, pp. 246–260, 2017, doi: 10.1109/TNSM.2017.2668363.

[22] A. Zangari, M. Marcuzzo, M. Schiavinato, A. Gasparetto, and A. Albarelli, "Ticket automation: An insight into current research with applications to multi-level classification scenarios," *Expert Systems with Applications*, vol. 225, p. 119984, Sep. 2023, doi: 10.1016/j.eswa.2023.119984.

[23] J. Xu, L. Tang, and T. Li, "System situation ticket identification using SVMs ensemble," *Expert Systems with Applications*, vol. 60, pp. 130–140, 2016, doi: 10.1016/j.eswa.2016.04.017.

[24] E. E. Jan, K. Y. Chen, and T. Ide, "Probabilistic text analytics framework for information technology service desk tickets," *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*, pp. 870–873, 2015, doi: 10.1109/INM.2015.7140397.

[25] Y. Li and T. H. Li, "A method of effort estimation for incident tickets in IT services," in *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2013*, Jul. 2013, pp. 311–316, doi: 10.1109/SOLI.2013.6611432.

[26] G. B. Dasgupta, T. K. Nayak, A. R. Akula, S. Agarwal, and S. J. Nadgowda, "Towards auto-remediation in services delivery: Context-based classification of noisy and unstructured tickets," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8831, pp. 478–485, 2014, doi: 10.1007/978-3-662-45391-9_39.

[27] C. Kadar, D. Wiesmann, J. Iria, D. Husemann, and M. Lucic, "Automatic classification of change requests for improved IT service quality," *Proceedings - 2011 Annual SRII Global Conference, SRII 2011*, pp. 430–439, 2011, doi: 10.1109/SRII.2011.95.

[28] S. Agarwal, R. Sindhgatta, and B. Sengupta, "SmartDispatch: Enabling efficient ticket dispatch in an IT service environment," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2012, pp. 1393–1401, doi: 10.1145/2339530.2339744.

[29] R. Potharaju, N. Jain, and C. Nita-Rotaru, "Juggling the jigsaw: Towards automated problem inference from network trouble tickets," *Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2013*, pp. 127–141, 2013.

[30] V. Shimpi, M. Natu, V. Sadaphal, and V. Kulkarni, "Problem Identification by Mining Trouble Tickets," *COMAD '14 Proceedings of the 20th International Conference on Management of Data*, pp. 76–86, 2014.

[31] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, Mar. 2002, doi: 10.1145/505282.505283.

[32] Y. H. Li and A. K. Jain, "Classification of text documents," *Computer Journal*, vol. 41, no. 8, pp. 543–545, Aug. 1998, doi: 10.1093/comjnl/41.8.537.

[33] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text classification using machine learning techniques," *WSEAS Transactions on Computers*, vol. 4, no. 8, pp. 966–974, 2005.

[34] B. Baharudin, L. H. Lee, and K. Khan, "A review of machine learning algorithms for text-documents classification," *Journal of Advances in Information Technology*, vol. 1, no. 1, Feb. 2010, doi: 10.4304/jait.1.1.4-20.

[35] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets : A review," *Science*, vol. 30, no. 1, pp. 25–36, 2006, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.9248&rep=rep1&type=pdf.

[36] L. S. Larkey and W. B. Croft, "Combining classifiers in text categorization," in *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, 1996, pp. 289–297, doi: 10.1145/243199.243276.

[37] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear Estimation and Classification*, Springer New York, 2003, pp. 149–171.

[38] L. Moreira-Matias, J. Mendes-Moreira, J. Gama, and P. Brazdil, "Text categorization using an ensemble classifier based on a mean co-association matrix," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7376 LNAI, Springer Berlin Heidelberg, 2012, pp. 525–539.

[39] R. Narayan, M. Roy, and S. Dash, "Ensemble based hybrid machine learning approach for sentiment classification- A review," *International Journal of Computer Applications*, vol. 146, no. 6, pp. 31–36, Jul. 2016, doi: 10.5120/ijca2016910813.

[40] P. Zicari, G. Folino, M. Guarascio, and L. Pontieri, "Combining deep ensemble learning and explanation for intelligent ticket management," *Expert Systems with Applications*, vol. 206, p. 117815, Nov. 2022, doi: 10.1016/j.eswa.2022.117815.

[41] C. C. Chen, C. H. Chang, and K. L. Hsiao, "Exploring the factors of using mobile ticketing applications: Perspectives from innovation resistance theory," *Journal of Retailing and Consumer Services*, vol. 67, p. 102974, Jul. 2022, doi: 10.1016/j.jretconser.2022.102974.

[42] J. Schad, R. Sambasivan, and C. Woodward, "Predicting help desk ticket reassignments with graph convolutional networks," *Machine Learning with Applications*, vol. 7, p. 100237, Mar. 2022, doi: 10.1016/j.mlwa.2021.100237.
[43] H. Zhang, Q. Zheng, B. Dong, and B. Feng, "A financial ticket image intelligent recognition system based on deep learning," *Knowledge-Based Systems*, vol. 222, p. 106955, Jun. 2021, doi: 10.1016/j.knosys.2021.106955.
[44] E. Stavinova, P. Chunaev, and K. Bochenina, "Forecasting railway ticket dynamic price with Google Trends open data," *Procedia Computer Science*, vol. 193, pp. 333–342, 2021, doi: 10.1016/j.procs.2021.10.034.
[45] L. Nandanwar *et al.*, "DCT-phase statistics for forged IMEI numbers and air ticket detection," *Expert Systems with Applications*, vol. 164, p. 114014, Feb. 2021, doi: 10.1016/j.eswa.2020.114014.
[46] J. Xu, J. Mu, and G. Chen, "A multi-view similarity measure framework for trouble ticket mining," *Data and Knowledge Engineering*, vol. 127, p. 101800, May 2020, doi: 10.1016/j.datak.2020.101800.
[47] T. Wang, P. Wu, Q. Ge, and Z. Ning, "Ticket prices and revenue levels of tourist attractions in China: Spatial differentiation between prefectural units," *Tourism Management*, vol. 83, p. 104214, Apr. 2021, doi: 10.1016/j.tourman.2020.104214.
[48] I. Koc and E. Arslan, "Dynamic ticket pricing of airlines using variant batch size interpretable multi-variable long short-term memory," *Expert Systems with Applications*, vol. 175, p. 114794, Aug. 2021, doi: 10.1016/j.eswa.2021.114794.
[49] J. Xu and R. He, "Expert recommendation for trouble ticket routing," *Data and Knowledge Engineering*, vol. 116, pp. 205–218, Jul. 2018, doi: 10.1016/j.datak.2018.06.004.
[50] Q. He and X. Zheng, "Price discrimination across different ticket distribution channels: Evidence from the US-china flight market," *China Economic Review*, vol. 61, p. 101236, Jun. 2020, doi: 10.1016/j.chieco.2018.10.003.

BIOGRAPHIES OF AUTHORS

Mudragada.Venkata Subbarao Researcg Scholar in Department of CS & SE, AUCE(A), Visakhapatnam, Andhra Pradesh, india. His research interests are in the field of Big Data, Cloud Computing, Machine Learning, Artificial Intelligence. He Published papers in national and international journals and also in international conferences including Springer. He has successfully guided a good number of the undergraduate and postgraduate thesis 25 Present he is doing research Department of CS & SE, AUCE(A), Visakhapatnam, Andhra Pradesh, india. He can be contacted at email: subbaraomudragada@gmail.com.Google Scholar: M. VENKATA SUBBARAO, Scopus: Venkata Subbarao, M, ORCID:

Kasukurthi Venkatarao He currently working as professor in Department of CS & SE, AUCE(A), Visakhapatnam, Andhra Pradesh, india. His research interests are in the field of Big Data, Cloud Computing, Internet of things, Artificial Intelligence, Computer Networks. He Published papers in national and international journals and also in international conferences including Springer, IEEE, Elesiver. He has successfully guided a good number of the undergraduate and postgraduate thesis 30. Present he is working in Department of CS & SE, AUCE(A) Visakhapatnam, Andhra Pradesh, India. He can be contacted at email: professor_venkat@yahoo.com.

Dr Suresh Chittineni He currently working as professor in Department of Computer Science & Engineering, Gitam University, Andhrapradesh, Visakhapatnam, India. His research interests are in the field of Big Data, Cloud Computing, Internet of things, Artificial Intelligence. He Published papers in national and international journals and in international conferences including Springer, IEEE, Elsevier. He has successfully guided a good number of the undergraduate and postgraduate thesis 17. Present he is working Prefessor in Department of Computer Science & Engineering, Gitam University, Vizag, India. He can be contacted at email: schittin@gitam.edu.

Kompella, Subhadra She currently working as Associate Professor in Department of Computer Science & Engineering, Gitam University, Andhrapradesh, Visakhapatnam, India. Her research interests are in the field of Text Mining, and Deep learning. She published papers in national and international journals and in international conferences including Springer, IEEE, Elsevier. She has successfully guided a good number of undergraduate and postgraduate theses 12. Present she is working as a Associate Prefessor in Department of Computer Science & Engineering, Gitam University, Vizag, India. She can be contacted at email: skompell@gitam.edu.