

WEB Traffic Management by Techniques with Support Vector Machine, Naïve Bayes and Discriminant Analysis

Ivelina Balabanova

Dept. of Communications Equipment
and Technologies
Technical University of Gabrovo
Gabrovo, Bulgaria
ivstoeva@abv.bg

Desislava Petrova

Dept. of Management
Technical University of Gabrovo
Gabrovo, Bulgaria
des_petrova@abv.bg

Georgi Georgiev

Dept. of Communications Equipment
and Technologies
Technical University of Gabrovo
Gabrovo, Bulgaria
givanow@abv.bg

Abstract—Effective diagnostics of WEB services distribution is an important prerequisite for adequate traffic load service planning in information and communication infrastructures in serving corporate clients. Network administrators play a key role in specification and transmission medium metrics analysis in web traffic management. The paper proposes a Machine Learning approach for WEB traffic streams diagnostics and categorization of corporate business clients in urban areas. The approach envisages a pre-processing phase in network traffic analysis procedures, incorporating the following specific traffic metrics: Flows, IPv4 Packet size, IPv4 Packet size distribution, and Mean Transition Rate. The second stage of the research integrates analytical tools based on Support Vector Machine (SVM), Naïve Bayes (NB) algorithm, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA). The effectiveness of a system comprising C-SVM and Nu-SVM classifiers was evaluated using different sets of Kernel functions - Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid types were applied. An advantage for Nu-SVM WEB traffic classification processes was established, where a maximum threshold recognition of 100.00% was achieved for the target classification groups defined. NB classifier training processes were conducted by Normal, Lognormal, Gamma and Poisson distributions applied to the input datasets. The highest classification accuracy of 100.00% was obtained for the Normal distribution, whereas when training the NB classifier with the Lognormal data distribution a significant reduction in this criterion below the 36.00% threshold was observed. Training and verification series of procedures were performed for Discriminant classifiers, respectively LDA, Diagonal Linear Discriminant Analysis (DLDA), Pseudo-Linear Discriminant Analysis (PLDA), QDA, Diagonal Quadratic Discriminant Analysis (DQDA) and Pseudo-Quadratic Discriminant Analysis (PQDA). In performance evaluation for the defined Linear and Quadratic types of DA classifiers, the highest recognition threshold of 90.00% was

achieved in the WEB traffic streams analyzed in the case of PLDA classification procedures.

Keywords—network monitoring; WEB traffic, urban area, C-SVM, Nu-SVM, Naïve Bayes; linear DA, quadratic DA.

I. INTRODUCTION

Ensuring adequate quality of service for private and corporate clients in WEB Traffic and Network Resources parametrization encompasses a number of specific activities. Some of the most important ones that can be mentioned are:

- Network traffic classification and prediction;
- Effective traffic routing and network security;
- Network performance and data integrity support;
- Error control in transmission media surcharge, [1-4] etc.

The implementation of analytical approaches for network diagnostics will further expand the spectrum of developers' capabilities for modeling and evaluating:

- Communication channel specification, noise control in transmitter and receiver services;
- Effective traffic load monitoring in structural network resources;
- Local coverage geographical areas assessment;
- Reliability and services prioritization in broadband Internet access planning [5-7].

More specifically, tools based on Artificial Intelligence, Machine Learning and Deep Learning concepts have been used to evaluate various processes regarding the

Online ISSN 3044-7771

<https://doi.org/10.68302/std2026.vol3.122>

© 2026 The Author(s). Published by Green Future Technologies.

This is an open-access article under the [Creative Commons Attribution 4.0 International License](#).

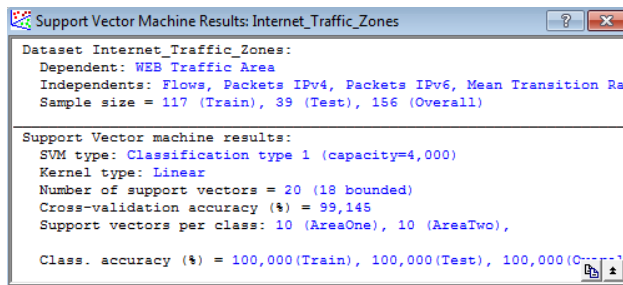
effectiveness of the communication infrastructure, as follows:

- Channel estimation: DNN (Deep Neural Networks), LSTM (Long-Short Term Memory) – a variant of RNN (Recurrent Neural Network);
- Channel prediction: Deep Neural Networks and CNN (Convolutional Neural Networks);
- Channel coding: Deep Neural Networks;
- Link adaptation: SVM (Support Vector Machine), k-NN (k – Nearest Neighbors);
- Reflecting intelligent surfaces (RIS): Deep Neural Networks;
- Spectrum sensing: k-Means, GMM (Gaussian Mixture Models), Support Vector Machine, k-Nearest Neighbors, Convolutional Neural Networks [8-12].

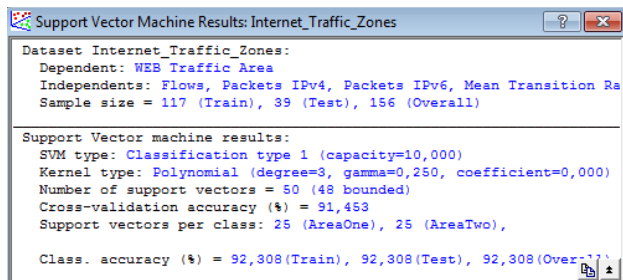
The paper proposes an approach based on Machine Learning methods and algorithms Support Vector Machine, Naïve Bayes and Discriminant Analysis for diagnosing the distribution of WEB services within active time zone in urban area for private and corporate clients.

II. SYNTHESIS OF SVM MODELS FOR WEB TRAFFIC AREA IDENTIFICATION IN STATISTICA

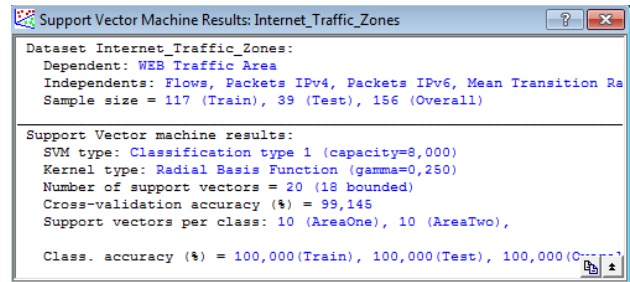
The first phase of the research is aimed at assessing the applicability of the SVM method on the defined task of diagnosing regions of consumption of accessible WEB content in urban environments. Two categories of SVM classifiers are adapted, respectively C-SVM and Nu-SVM. The evaluation of the functionality of the first group of classifiers was carried out under the following conditions: "Linear"; "Polynomial"; "RBF" and "Sigmoid" Kernel functions, shown in Fig. 1.



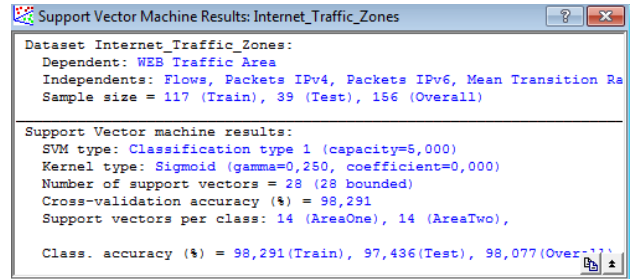
a)



b)



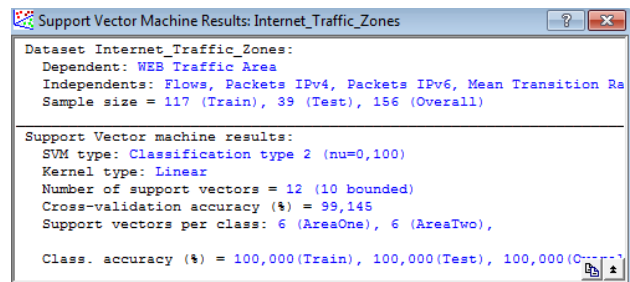
c)



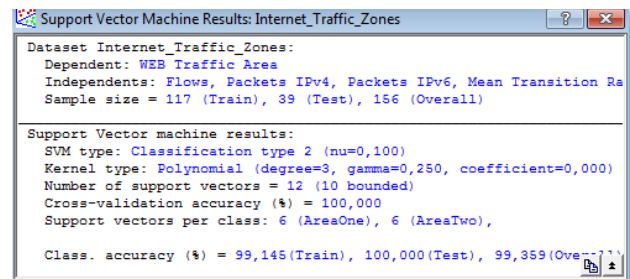
d)

Fig. 1. C-SVM classifiers for WEB traffic area recognition.

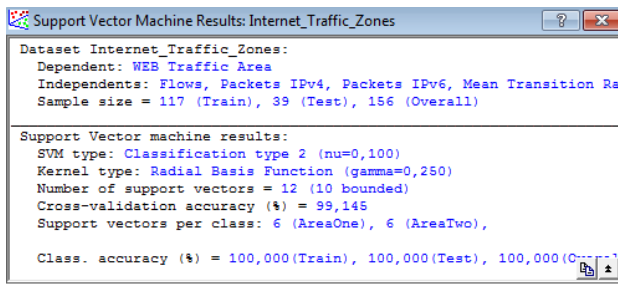
The necessary "Support Vectors" for constructing the "Hypersurface" separating the target output groups in the classification space, respectively "20", "50", "20" and "28", were analytically determined. The lowest final accuracy of 92.308 % was established for the C-SVM model with the applied Kernel function "Polynomial". A complete correct recognition of 100.0 % of the standards participating in the training and test datasets has been achieved, matching for the cases of "Linear" and "RBF", while in "Sigmoid" Kernel type final reported accuracy is 98.077 %.



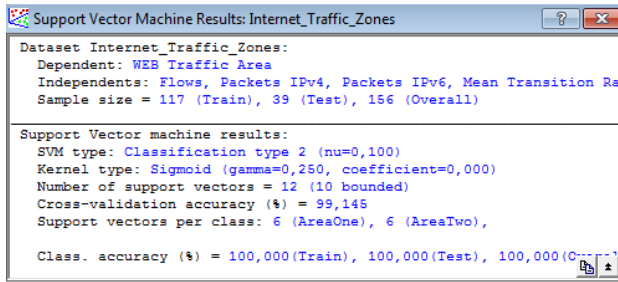
a)



b)



c)



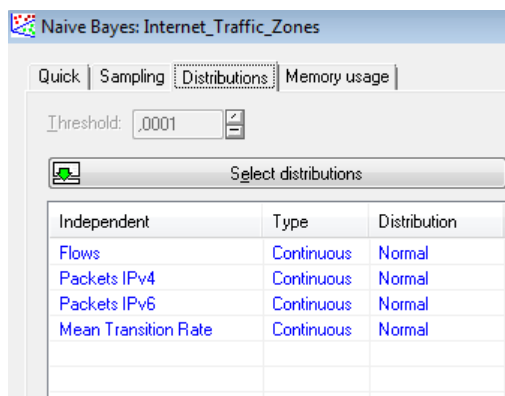
d)

Fig. 2. Nu-SVM classification models in WEB traffic area monitoring.

The possibility of adapting Nu-SVM classification models with respect to analogous types of Kernel functions was evaluated with some advantage, shown in Fig. 2. The maximum recognition threshold of 100.0% was reached for “Linear”, “Radial Basis Function” and “Sigmoid” Kernel functions. The final accuracy obtained with the Nu-SVM model using the “Polynomial” type set is 99.359 %.

III. NAÏVE BAYES CLASSIFICATION MODEL ANALYSIS IN WEB TRAFFIC MONITORING PROCEDURES IN STATISTICA

The second phase of the research consists of evaluating the suitability of the Naïve Bayes algorithm for four different types of distributions of variables in input datasets, related to the task of recognizing Internet usage zones and traffic load of network resources for corporate clients. The following types of data distributions were applied, respectively: “Normal”; “Lognormal”; “Gamma” and “Poisson”, given in Fig. 3.a), Fig. 4.a), Fig. 5.a), and Fig. 6.a).



a)

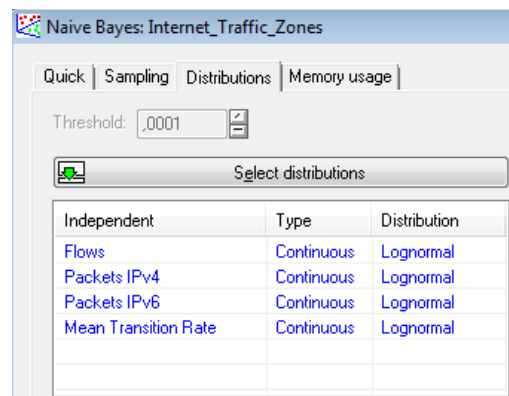
Naive Bayes model, WEB Traffic Area (Internet_Traffic_Zones)					
Class Name	Prior values	Flows Mean	Flows Standard deviation	Packets IPv4 Mean	Packets IPv4 Standard deviation
AreaOne	0.307692	125.9167	200.3643	123.2222	180.6921
AreaTwo	0.692308	200.5679	973.7235	179.9012	819.8901

b)

Classification summary (Naive Bayes Classifier), WEB					
Class Name	Total	Correct	Incorrect	Correct(%)	Incorrect(%)
AreaOne	14	14	0	100.0000	0.00
AreaTwo	25	25	0	100.0000	0.00

c)

Fig. 3. NB classifier in Normal distribution – a) distribution type, b) statistical metrics and c) confusion matrix.



a)

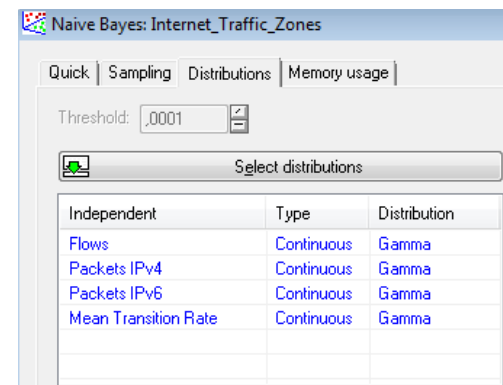
Naive Bayes model, WEB Traffic Area (Internet_Traffic_Zones)					
Class Name	Prior values	Flows Scale	Flows Shape	Packets IPv4 Scale	Packets IPv4 Shape
AreaOne	0.307692	66.99912	1.123335	69.42450	1.071214
AreaTwo	0.692308	40.46357	1.789274	38.55687	1.755149

b)

Classification summary (Naive Bayes Classifier), WEB					
Class Name	Total	Correct	Incorrect	Correct(%)	Incorrect(%)
AreaOne	39	14	25	35.89744	64.10256
AreaTwo	0	0	0	0.00000	0.00000

c)

Fig. 4. NB classification model in Lognormal distribution of input variables – a) distribution type, b) statistical metrics and c) confusion matrix.



a)

Naive Bayes model, WEB Traffic Area (Internet Traffic Zones)					
Class Name	Prior values	Flows Scale	Flows Shape	Packets IPv4 Scale	Packets IPv4 Shape
AreaOne	0.307692	318.829	0.394935	264.965	0.465050
AreaTwo	0.692308	4727.264	0.042428	3736.605	0.048146

b)

Classification summary (Naive Bayes Classifier), WEB					
Class Name	Total	Correct	Incorrect	Correct(%)	Incorrect(%)
AreaOne	0	0	0	0.00000	0.00000
AreaTwo	39	25	14	64.10256	35.89744

c)

Fig. 5. NB classifier at Gamma data distribution – a) distribution type, b) statistical metrics and c) confusion matrix.

Statistical metrics “Mean” and “Standard Deviation” were calculated for each input feature, as follows “Flows”, “IPv4 Packet size”, “Packets in IPv6” and “Mean Transition Rate”. The analytical indicators extracted from the quantitative analysis are given in Fig 3.b), Fig. 4.b), Fig. 5.b), and Fig. 6.b).

Naive Bayes: Internet_Traffic_Zones		
Quick	Sampling	Distributions
Threshold: .0001		
Select distributions		
Independent	Type	Distribution
Flows	Continuous	Poisson
Packets IPv4	Continuous	Poisson
Packets IPv6	Continuous	Poisson
Mean Transition Rate	Continuous	Poisson

a)

Naive Bayes model, WEB Traffic Area (Internet Traffic)				
Class Name	Prior values	Flows Shape	Packets IPv4 Shape	Packets IPv6 Shape
AreaOne	0.307692	125.9167	123.2222	123.6667
AreaTwo	0.692308	200.5679	179.9012	182.0741

b)

Classification summary (Naive Bayes Classifier), WEB					
Class Name	Total	Correct	Incorrect	Correct(%)	Incorrect(%)
AreaOne	17	14	3	82.3529	17.64706
AreaTwo	22	22	0	100.0000	0.00000

c)

Fig. 6. NB classification model at Poisson distribution of input variables -a) distribution type, b) statistical metrics and c) confusion matrix.

According to the “Normal” distribution of informative features, correct classification of the informative standards from both source groups was reported. With “Lognormal” data distribution, 25 samples from “Class № 2” were incorrectly assigned to samples from “Class № 1”. In the case of “Gamma” distribution, 14 samples included in

“Class № 1” were characterized by incorrectly determined membership to the composition of “Class № 2”. Regarding the last applied input dataset distribution, 3 samples from the structure of “Class № 2” were incorrectly assigned to the composition of the first output group. The obtained evaluation results show the highest degree of applicability to a standard “Normal” data distribution, confirmed in Fig. 3.c), Fig. 4.c), Fig. 5.c), and Fig. 6.c).

IV. EFFICIENCY ANALYSIS OF LINEAR AND QUADRATIC DISCRIMINANT CLASSIFIERS FOR WEB TRAFFIC DIAGNOSTICS IN MATLAB

The third phase of the research is based on evaluating the effectiveness of the created “Linear” and “Quadratic” Discriminant classification models in WEB Traffic monitoring procedures. The “Linear”, “Diagonal Linear”, “Pseudo-Linear”, “Quadratic”, “Diagonal Quadratic” and „Pseudo-Quadratic“ were successively evaluated. Table 1 summarizes the results for the studied discriminant models, confirming good and comparable achieved efficiency above 85.0 % levels.

TABLE I. ASSESSMENT METRICS IN EXAMINING THE QUALITY OF DISCRIMINANT CLASSIFIERS IN WEB TRAFFIC MONITORING

№	Evaluation Techniques		
	Type of Discriminant Classifier	Cross-Validation Loss	Cross-Validation Accuracy, %
1.	Linear	0.1026	89.47
2.	Diagonal Linear	0.1218	87.82
3.	Pseudo-Linear	0.0962	90.38
4.	Quadratic	0.1154	88.46
5.	Diagonal Quadratic	0.1218	87.82
6.	Pseudo-Quadratic	0.1026	89.74

Variables - DAModel			
DAModel			
Tx1 ClassificationDiscriminant			
Property	Value	Min	Max
XCentered	156x4 double	-683.1964	899.2308
Y	156x1 cell		
X	156x4 double	2.4000	1260
W	156x1 double	0.0064	0.0064
ModelParameters	1x1 classreg.learn...		
NumObservations	156	156	156
PredictorNames	1x4 cell		
CategoricalPredictors	[]		
ResponseName	'Y'		
ClassNames	2x1 cell		
Prior	[0.5000 0.5000]	0.5000	0.5000
Cost	[0 1; 1 0]	0	1
ScoreTransform	'none'		
DiscrimType	'pseudoLinear'		
Gamma	0	0	0
Delta	0	0	0
Coeffs	2x2 struct		
Mu	[9.2324 691.0364 ...	9.2324	861.9615
Sigma	4x4 double	-3.4249e...	8.2681e+04

a)

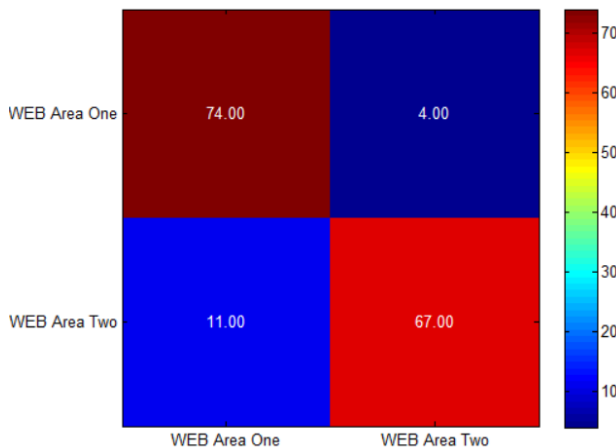
Variables - DAModel			
DAModel			
1x1 ClassificationDiscriminant			
Property	Value	Min	Max
XCentered	156x4 double	-683.1964	899.2308
Y	156x1 cell		
X	156x4 double	2.4000	1260
W	156x1 double	0.0064	0.0064
ModelParameters	1x1 classreg.learn...		
NumObservations	156	156	156
PredictorNames	1x4 cell		
CategoricalPredictors	[]		
ResponseName	'Y'		
ClassNames	2x1 cell		
Prior	[0.5000 0.5000]	0.5000	0.5000
Cost	[0 1; 1 0]	0	1
ScoreTransform	'none'		
DiscrimType	'diagQuadratic'		
Gamma	1	1	1
Delta	0	0	0
Coeffs	2x2 struct		
Mu	[9.2324 691.0364 ...]	9.2324	861.9615
Sigma	1x4x2 double	20.2860	9.6436e+04

b)

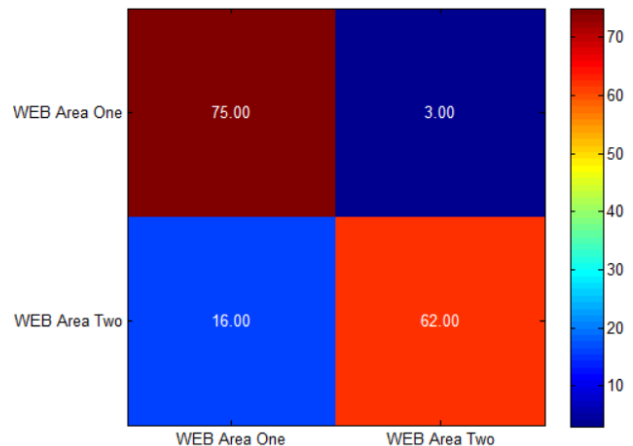
Fig. 7. Statistical indicators computed in a) “Pseudo-Linear” and b) “Diagonal Quadratic” model assessment.

The lowest classification accuracy of 87.82 % was reported for the cases of „Diagonal Linear“ and „Diagonal Quadratic“ discriminant type. The best adequacy to the specificity of the classification task was observed for „Pseudo-Linear“ model, expressed with a quantitative assessment of 90.38 %.

The set of specific statistical metrics computed in the course of studying the constructed Discriminant Classifiers shows the highest efficiency obtained in the case of “Pseudo-Linear” (Fig. 7.a) and the lowest efficiency achieved for the case of “Diagonal Quadratic” model, presented in Fig. 7.a). The set of specific variables includes “Gamma”, “Delta”, “Sigma”, “Mu”, etc. The matrix of correct membership defined and misclassification on Fig. 8 shows a total of 141 and 137 traffic samples with correctly defined membership in verification procedures, out of a total of 156 informative observations.



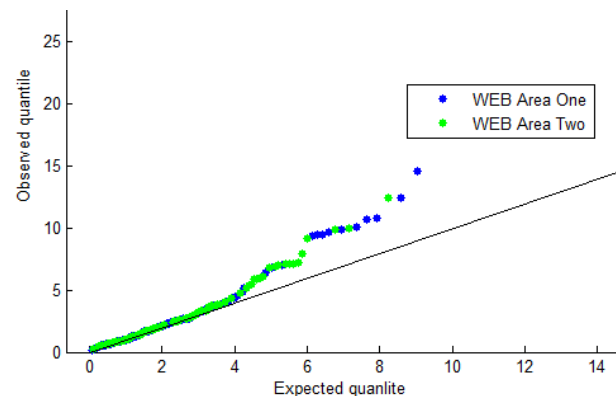
a)



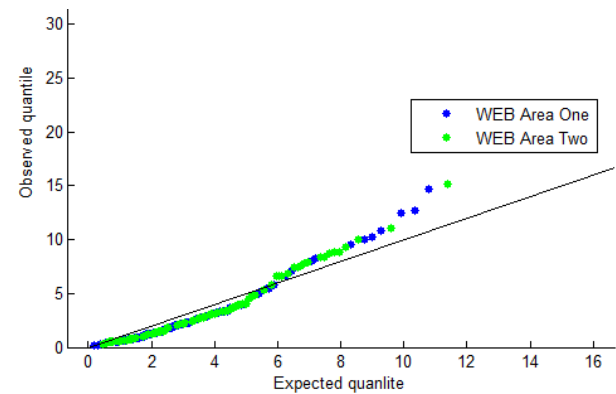
b)

Fig. 8. Confusion matrices in “Pseudo-Linear” and “Diagonal Quadratic” model assessment process.

In connection with confirming the correctness of the application of the analytical tool, “Quantile-Quantile” plots were constructed to compare two probability distributions by displaying and evaluating the behavior of “Observed” and “Executed” quantiles. The resulting Q-Q diagrams for the cases of “Pseudo-Linear” and “Diagonal Quadratic” discriminant models are shown in Fig. 9.



a)



b)

Fig. 9. Q-Q diagrams for “Pseudo-Linear” and “Diagonal Quadratic” discriminant model verification procedures.

For a significant share of statistical analyzing, the data sample possesses a "Gaussian distribution", which is a positive indication of the correctness of the use of the selected mathematical apparatus.

V. CONCLUSION

Effective diagnostics of network traffic with Machine Learning tools will significantly improve the efficiency of information services, Internet service providers and network operators. The information datasets analyzed covers complex metrics for registered traffic in relation to 200 companies for two geographic areas in the central part of the city of Chicago in active time zones 9.00 to 12 a.m. / 1.00 to 4.00 p.m. The total data is divided in a basic ratio of 2:1, respectively for training / testing procedures and evaluation through validation models for verification. The conducted research represents one of the main technical directions in this field. Future technical prospects for upgrading the scope of the considered issues can be divided into the following categories:

- Integrating classification and predictive analysis in terms of Cybersecurity management when ensuring access to electronic resources and services;
- Introduction of Design of experiment concept and non-linear minimization by optimization methods and algorithms related to complex performance indices of transmission media and traffic load evaluation of network servers, distribution units and system modules in communication and information systems.

ACKNOWLEDGMENT

The research was funded by Project "Quality of Service and Cybersecurity on Network Diagnostics in Communications by Artificial Intelligence" – Technical University of Gabrovo, Bulgaria.

REFERENCES

- [1] A. Shahraki, M. Abbasi, A. Taherkord and A. Jercut, "Active learning for network traffic classification: A technical study," IEEE Transactions on Cognitive Communications and Networking Journal, vol. 8, no. 1, p. 422+, October 2021. Available: DOI: [10.1109/TCCN.2021.3119062](https://doi.org/10.1109/TCCN.2021.3119062).
- [2] T. Vashishth, V. Sharma, B. Kumar and S. Chaudhary, "Artificial intelligence-enabled traffic optimization: A comprehensive survey," Industrial Engineering Journal, vol. 52, no. 5, p. 26+, May 2023.
- [3] A. Guo and C. Yuan, "Network intelligent control and traffic optimization based on SDN and artificial intelligence," Electronics Journal, vol. 10, no. 6, p. 2+, March 2021. Available: <https://doi.org/10.3390/electronics10060700>.
- [4] M. Latah and L. Toker, "Application of artificial intelligence to software defined networking: A survey," Indian Journal of Science and Technology, vol. 9, no. 46, p. 3+, November 2016. Available: DOI: [10.17485/ijst/2016/v9i44/89812](https://doi.org/10.17485/ijst/2016/v9i44/89812).
- [5] H. Shein, "Signal processing technology of mobile communication system based on artificial intelligence", Procedia Computer Science Journal, vol. 261, p. 1284+, December 2025. Available: <https://doi.org/10.1016/j.procs.2025.05.003>.
- [6] T. Singasani, M. Shaik, C. Gangajaliya, S. Ogety, V. B. Nallam and S. Katta, "AI-driven signal processing: Improving communication systems with machine learning-based noise reduction," in First International Conference on Radio Frequency Communications and Networks India, 2025, pp. 1-6.
- [7] Z. Bai, "Research on Application of Artificial intelligence in communication network," IOP Publishing Journal of Physics, vol. 2209, no. 1, p. 3+, February 2022. Available: DOI: [10.1088/1742-6596/2209/1/012014](https://doi.org/10.1088/1742-6596/2209/1/012014).
- [8] G. Fontanesi et al., "Artificial intelligence for satellite communication: A survey," IEEE Communications Surveys & Tutorials, vol. 28, no. 1, p. 1381+, February 2025. Available: DOI: [10.1109/COMST.2025.3534617](https://doi.org/10.1109/COMST.2025.3534617).
- [9] A. Thakur, "Artificial Intelligence (AI) in Information and Communication Technology (ICT): An Overview," International Journal of Research and Analysis in Science and Engineering, vol. 1, no. 3, p. 34+, May 2021.
- [10] K. Naseer Qureshi and T. Newe, Artificial Intelligence of Things: New Standards, Technologies and Communication Systems. NY: CRC Press Taylor & Francis Group, 2024.
- [11] K. Koufos, K. Haloju, M. Dianati, M. Higgins, J. Elmirghani, M. Imran and R. Tafazolli, "Trends in intelligent communication systems: Review of standards, major research projects, and identification of research gaps," MDPI Journal of Sensors and Actuator Networks, vol. 10, no. 4, p. 23+, October 2021. Available: <https://doi.org/10.3390/jsan10040060>.
- [12] Y. Mao, A. Pranolo, L. Hernandez, A. Wibawa and Z. Nuryana, "Artificial intelligence in mobile communication: A Survey," IOP Conference Series Journal: Materials Science and Engineering, vol. 1212, no. 1, p. 2+, January 2022. Available: DOI: [10.1088/1757-899X/1212/1/012046](https://doi.org/10.1088/1757-899X/1212/1/012046).