Unsupervised Learning Approach for Network Traffic Classification

Mario Bou Abboud*, Oumaya Baala*, Maroua Drissi†, Sylvain Allio†
* *UTBM, CNRS, Institut FEMTO-ST*
F-90010 Belfort cedex, France
† *Orange Labs, France*
mario.bouabboud@orange.com, oumaya.baala@utbm.fr,
maroua1.drissi@orange.com, sylvain.allio@orange.com

***Abstract*—The landscape of network management has undergone significant transformation with the advent of diverse Internet applications, smart devices, and the shift towards software-defined networks (SDN). This evolution has amplified the complexities of managing and measuring network traffic, necessitating more sophisticated and dynamic traffic classification methods to maintain optimal network performance and ensure user Quality-of-Experience (QoE). This paper presents a novel approach to network traffic classification, leveraging the capabilities of Gaussian Mixture Models (GMM) to classify network traffic based on user behavior patterns and temporal data. Our methodology distinctly categorizes network traffic into business or pleasure-oriented activities by analyzing various features such as the number of connected users, traffic volume, the day of the week, and the time of day. This classification is crucial not only for traffic management but also for understanding evolving network usage patterns, which are vital for ensuring robust network operations and efficient resource allocation.**

***Index Terms*—Network Traffic Classification, Urban Mobility Patterns, Gaussian Mixture Models, Machine Learning in Networking, Quality of Service (QoS), Smart City Applications, Real-time Data Analysis, Dynamic Bandwidth Allocation, Network Resource Management**

I. Introduction

In the evolving landscape of network management, the rapid proliferation of smart devices has introduced a new level of complexity, prompting network operators to explore novel management concepts. This dynamic and increasingly heterogeneous network environment poses a significant challenge in designing architectures that can efficiently handle diversity while maximizing resource utilization [14]. In response, innovative solutions such as Network Slicing (NS) and Machine Learning (ML) have emerged [15]. These technologies, when properly applied, pave the way for an autonomic and intelligent network resource management framework, essential for meeting the demanding QoS requirements of 5G and beyond. ML, recognized for its ability to solve complex problems without explicit programming, models and learns underlying behaviors using training datasets. Its efficiency and effectiveness have been demonstrated across various domains, including network management [8]. Concurrently, NS, offering the network as a service (NaaS), allows for the partitioning of network infrastructure into isolated slices, each with dedicated resources and performance requirements [6]. NS's adaptability to a wide array of services, support for multi-tenancy, and efficient resource utilization make it an ideal solution for the evolving needs of 5G networks [11]. However, the constantly changing network behaviors driven by factors such as user mobility, location, and social events make the explicit definition of network states a daunting task [2]. This challenge is compounded when allocating resources to network slices, especially considering the growing volume of network traffic and the diversity of new applications. There is a pressing need for a flexible, instantaneous, and traffic behavior-dependent approach to network slice creation and management [17]. Intelligent traffic management, which can understand the behavior of connected smart devices and applications, is vital for monitoring network slice performance and optimizing network resources. In this context, ML emerges as a powerful tool, capable of analyzing large datasets to discern useful patterns within a reasonable timeframe. This ability to handle complex problems and analyze vast datasets is integral for automating network tasks and enabling the transition towards self-configuring, self-healing, and self-optimizing networks [4]. Within ML, a distinction is made between supervised and unsupervised learning, with the former focused on mapping input features to output classes and the latter on discovering patterns in inputs without predefined classes. The efficacy of ML models is contingent on the quality and relevance of the data and features provided [19]. Increased data dimensionality can adversely affect algorithm performance and incur additional computational costs. Therefore, preprocessing steps, including feature selection, are crucial for preparing raw data for analysis. Our study introduces an ML-based solution for defining network slices. This architecture utilizes network statistics and an offline process for understanding network traffic patterns through clustering algorithms. Building upon these foundations, our research introduces a novel methodology for network traffic classification, employing GMM for data classification. Our approach, which considers both temporal and usage patterns, diverges from traditional models by analyzing attributes such as the number of connected users, traffic volume, day of the week, and time of the day. This enables us to categorize traffic into contextually relevant groups, like 'business' or 'pleasure' oriented traffic, offering insights into the evolving network use patterns. Our study aims to contribute

to the field by providing a nuanced understanding of network traffic, aiding in the optimization of network performance, and enhancing cybersecurity measures. This paper will explore the unique aspects of our approach in contrast to existing methodologies and discuss its potential impact on the future of network traffic classification and management.

II. Related Works

The realm of network traffic classification is rapidly evolving, marked by diverse and innovative approaches to address complex challenges in the field. In [5], a convolutional neural network is designed for efficient and explainable traffic classification. This model focuses on computational efficiency and the increasing need for explainable AI, incorporating a novel architecture with a residual block and a prototype layer for enhanced accuracy and insight into decision-making processes. Complementing this, [7] advanced the field with a federated semi-supervised learning approach that adeptly balances privacy concerns and the scarcity of labeled data, achieving a high accuracy rate of 97.81% on a public dataset. In parallel, [18] introduces Ulfar, a deep learning model utilizing a multi-scale feature attention mechanism, revolutionized the field by requiring only a single packet per flow for effective classification. Ulfar's flexibility in generating variable n-gram features and its applicability to diverse network environments positioned it as a frontrunner in network traffic analysis. Adding to these advancements, [9] leveraged Single Flow Time Series Analysis for encrypted traffic monitoring, introducing a set of 69 universal features derived from detailed time series analysis of network packets. Their method demonstrated superior performance in various classification scenarios across 15 publicly accessible datasets. In the context of SDN, [13] made significant strides in optimizing power consumption in data center networks. They introduced a Machine Learning-based online traffic classification system that efficiently balances power usage with Quality of Service, revealing the potential of integrating traffic classification into power consumption models. Similarly, [10] addressed the challenge of dataset imbalance in network traffic classification by employing a hybrid model integrating 1DCNN and BiLSTM, achieving notable success in minority class categorization. Innovations by [16] further diversified the landscape. Saqib et al. proposed a framework for traffic classification within programmable data planes, using sequential packet size information for classification, while [3] focused on classifying IoT and non-IoT traffic in home networks, revealing the challenges of adapting to new IoT device categories. Furthermore, [12] conducted a comprehensive evaluation of ensemble learning techniques, highlighting their superiority over conventional supervised algorithms in multi-label network traffic classification. A significant contribution came from [20], who prioritized Quality of Service awareness in network traffic classification. Their approach, utilizing a Long Short-Term Memory neural network-based Autoencoder, extracts QoS-aware features such as inter-APP similarity and intra-APP diversity. This method not only considers the source application but also acknowledges the varied QoS requirements, demonstrating its efficacy in network resource management. The paper [1] presents an approach to network traffic classification by leveraging unsupervised learning techniques, specifically GMM, to identify and verify network traffic patterns. This method stands out by not relying on pre-labeled data, offering a flexible solution to the challenges of encrypted and dynamic network environments. While it shares similarities with our work in the use of GMM for traffic classification, our research extends the application of these models by incorporating temporal and behavioral patterns to categorize traffic more granularly into business or pleasure-oriented activities. This distinction allows for a more nuanced understanding of network usage, contributing to more effective network management and resource allocation strategies. These diverse studies collectively contribute to the field of network traffic classification, each offering unique solutions to the challenges of effectively managing and securing network traffic.

III. Methodology/Model Architecture

A. Model Configuration

Gaussian Mixture Models are a probabilistic model for representing normally distributed subpopulations within an overall population. In the context of network traffic analysis, GMMs provide a sophisticated approach to clustering by modeling traffic data as a mixture of several Gaussian distributions. Each Gaussian, or 'component', in the mixture model represents a different group or cluster within the data, characterized by its own mean and variance. This makes GMMs particularly useful for identifying underlying patterns in data where clusters may overlap or have different variances, a common scenario in network traffic. One of the key strengths of GMMs lies in their flexibility. Unlike many other clustering techniques, GMMs do not require clusters to have a specific shape (such as spherical, as in k-means clustering). This allows GMMs to fit more complex cluster shapes and sizes, providing a more nuanced understanding of data distributions. Additionally, GMMs incorporate a probabilistic model, assigning each data point a probability of belonging to each cluster, rather than forcing a hard assignment. This aspect of GMMs is especially beneficial in scenarios where data points could reasonably belong to multiple clusters, allowing for a more accurate and interpretable clustering process. Moreover, GMMs facilitate the determination of the number of clusters present in the data. This is achieved through model selection criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), which balance model complexity with goodness of fit. Thus, GMMs offer a data-driven approach to determining the optimal number of clusters, which is a critical step in unsupervised learning tasks like network traffic classification. In applying GMM to network traffic classification, the model can discern intricate patterns within the traffic data, enabling the classification of network traffic into meaningful categories. This categorization is not merely based on superficial traffic characteristics but is grounded in the underlying statistical properties of the traffic

data. As a result, GMM-based analysis can reveal insights into network usage patterns and aid in optimizing network resource allocation, thereby enhancing overall network performance.

B. Mathematical Foundation of Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are employed to model complex datasets by assuming that the data are generated from a mixture of several Gaussian distributions with unknown parameters. Mathematically, the probability density function of a GMM is given by:

$$p(\mathbf{x}|\Theta) = \sum_{i=1}^{K} \pi_i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{1}$$

where $\mathbf{x}$ is a data point, $\Theta = \{\pi_1, \ldots, \pi_K, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\}$ represents the parameters of the mixture model, K is the number of Gaussian components in the mixture, π_i is the mixing coefficient of the i^{th} component, and $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the normal (Gaussian) distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. The mixing coefficients π_i satisfy the constraints $0 \leq \pi_i \leq 1$ and $\sum_{i=1}^{K} \pi_i = 1$. Each component in the mixture model captures a cluster in the data, with the Gaussian distribution representing the shape of the cluster. The mean $\boldsymbol{\mu}_i$ indicates the center of the cluster, while the covariance matrix $\boldsymbol{\Sigma}_i$ describes its shape and orientation. The mixing coefficient π_i reflects the relative proportion of data points belonging to the i^{th} component. The parameters of the GMM are typically estimated using the Expectation-Maximization (EM) algorithm, which iteratively optimizes the likelihood of the observed data. The EM algorithm consists of two steps: the Expectation (E) step, which computes the posterior probabilities of each component given the data, and the Maximization (M) step, which updates the parameters based on these probabilities. The EM algorithm alternates between these two steps until convergence. In the E step, the posterior probability of each component given a data point $\mathbf{x}$, also known as the responsibility $\gamma_i(\mathbf{x})$, is computed as:

$$\gamma_i(\mathbf{x}) = \frac{\pi_i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{2}$$

In the M step, the parameters are updated as follows:

$$\pi_i^{new} = \frac{1}{N} \sum_{n=1}^{N} \gamma_i(\mathbf{x}_n) \tag{3}$$

$$\boldsymbol{\mu}_i^{new} = \frac{\sum_{n=1}^{N} \gamma_i(\mathbf{x}_n) \mathbf{x}_n}{\sum_{n=1}^{N} \gamma_i(\mathbf{x}_n)} \tag{4}$$

$$\boldsymbol{\Sigma}_i^{new} = \frac{\sum_{n=1}^{N} \gamma_i(\mathbf{x}_n)(\mathbf{x}_n - \boldsymbol{\mu}_i^{new})(\mathbf{x}_n - \boldsymbol{\mu}_i^{new})^T}{\sum_{n=1}^{N} \gamma_i(\mathbf{x}_n)} \tag{5}$$

where N is the number of data points. Through the EM algorithm, GMM adapts its parameters to best fit the underlying structure of the data, effectively capturing the inherent clusters. This process results in a flexible model that can accommodate a wide range of data distributions, making GMMs a powerful tool for unsupervised learning in diverse applications such as network traffic analysis.

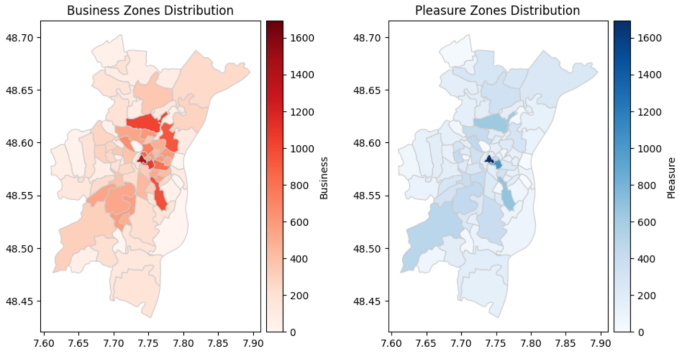

Fig. 1: Comparative Visualization of Business and Pleasure Zones (Venues) distribution in Strasbourg

IV. Experimental Results

A. Data Preparation

In the realm of unsupervised learning, the design and selection of features play a pivotal role in the effectiveness of the models. This is particularly evident in our dataset, which is a time series collected from the metropolitan area of Strasbourg using Orange's in-house tools. The dataset encapsulates the region divided into communes or zones, each characterized by a number of shops and services categorized as either Business or Pleasure. This division into zones provides a spatial granularity of 2000 different network cells, coupled with a temporal granularity of 15 minutes.

Key features of the dataset include:

- **Date and Time:** Days of the week are one-hot encoded to capture the cyclical nature of weekly activities.
- **Periods of the Day:** Time slots are divided into "00:00-09:00", "09:00-12:00", "12:00-14:00", "14:00-17:00", and "17:00-24:00", aiming to differentiate between business hours and regular hours, and they are also one hot encoded.
- **Number of Users Connected:** This feature indicates the volume of users active in each cell.
- **Amount of Traffic:** This measures the data usage in each cell, providing insights into the intensity of network activity.
- **Surface of the Zones:** The area of the zones to which each cell belongs, adding a spatial dimension to the analysis.
- **Business Venues:** Number of venues where business oriented Traffic is produced, showcased in Figure 1.
- **Pleasure Venues:** Number of venues where pleasure oriented Traffic is produced, showcased in Figure 1.

In our study, the terms 'business' and 'pleasure' serve as proxies to differentiate between network traffic for critical or

priority tasks and traffic for leisure or non-critical activities. It's essential to clarify that these classifications do not reflect the operational traffic management strategies of Orange but were devised solely for research purposes. For the venue classification within our network traffic study in Strasbourg, we leveraged data provided by INSEE, France's National Institute of Statistics and Economic Studies. INSEE's role in compiling comprehensive statistics on French economic and social structures is invaluable for our research. We utilized their detailed dataset, which includes the locations and general types of venues in the Strasbourg area, encompassing a broad spectrum from social services to tourism including social services, commerce, primary education, secondary education, higher education and training, para-medical functions, personal services, health services, sports leisure and culture, transport and tourism. This publicly available information is crucial for understanding and predicting the type of network traffic generated in different locales, as it mirrors the daily activities and functions likely to influence network demand at various times. For our experiment, we divided the 10 classes to 2 categories mainly business and pleasure. We categorized venues as 'business' if they are associated with activities that typically demand priority in network traffic handling. This includes venues related to social services, educational institutions at all levels including universities, and health-related facilities such as hospitals and clinics. On the other hand, 'pleasure' venues encompass areas associated with commerce, sports, leisure, culture, transport, and tourism—activities that, while important, could potentially tolerate network traffic delays without immediate adverse effects. We acknowledge that this binary categorization, albeit useful for our research framework, oversimplifies the complex spectrum of network usage. Therefore, we propose as a future research direction the development of a more nuanced, multi-tiered classification system that can capture the diverse and dynamic priorities of network traffic in a more granular manner. This dataset includes a variety of features designed to understand both where (spatial) and when (temporal) network activities happen. These features are crucial for our unsupervised learning methods, which we use to sort out and forecast the kinds of network traffic found in the Strasbourg metropolitan area.

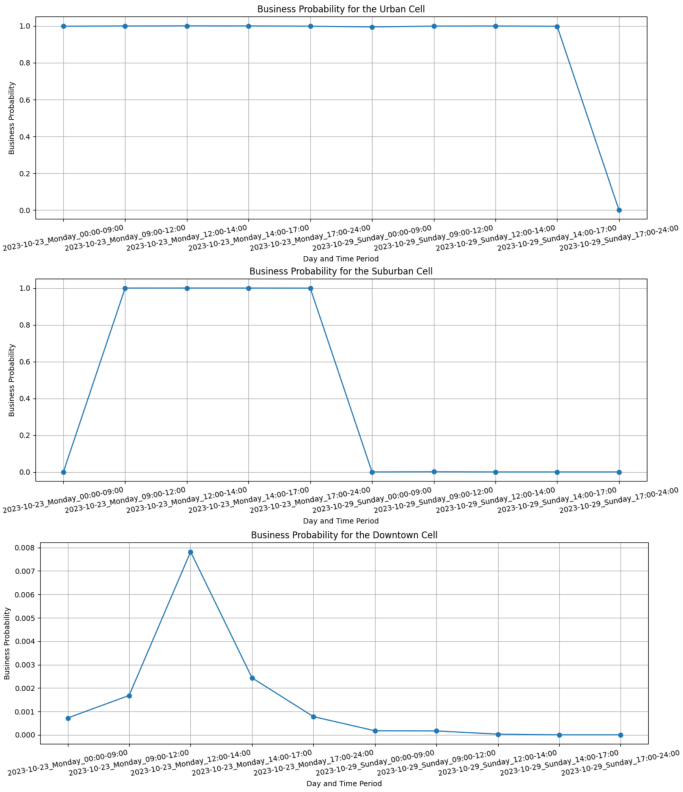

Fig. 2: Traffic Classification for three cellular sites situated in urban, downtown, and suburban areas on both a Monday and a Sunday.

B. Results

In the construction of our classification model, we employed the K-means clustering algorithm as a baseline method due to its simplicity and efficiency in handling large datasets. Remarkably, when compared to the outcomes from the Gaussian Mixture Model (GMM), we observed a high degree of alignment between the two techniques. Specifically, the K-means algorithm yielded results that were identical to the GMM classification in 72.2% of cases. However, the nuanced differences captured by GMM suggest that for a more refined and probabilistic understanding of venue classification, more sophisticated methods can offer additional insights. The experimental analysis of the traffic classification in Strasbourg's metropolitan area revealed insightful patterns of the two-cluster model, which, upon analysis, revealed a strong influence from the initial business and pleasure categorizations. This classification system was able to adapt to varying temporal and spatial dimensions, showcasing its ability to dynamically determine the nature of traffic—be it business or pleasure—based on a comprehensive set of factors. The three cells represented in Figure 2, each representing distinct urban dynamics – downtown cell, urban cell, and suburban cell – show how traffic can be categorized into business or pleasure-oriented, reflecting the inherent socio-economic activities of each area. Starting with urban cell, a business-centric zone, the graph shows a high probability of business traffic during the entire days of both Monday and Sunday then steeply declines post 17:00 on Sunday night, indicating a transition to pleasure-oriented activities. This pattern aligns with the typical business areas, where business activities dominate most of the days of the week. The second cell, located in the suburban outskirts, displays a balanced profile with a clear distinction between weekdays and weekends. On Monday, the traffic is more business-oriented, especially during conventional working hours, while on Sunday, the pleasure orientation takes precedence throughout the day, underscoring the weekend's recreational bias. In contrast, the downtown cell, a tourist hotspot, exhibits a relatively constant low probability of business traffic, with a slight spike in business-oriented

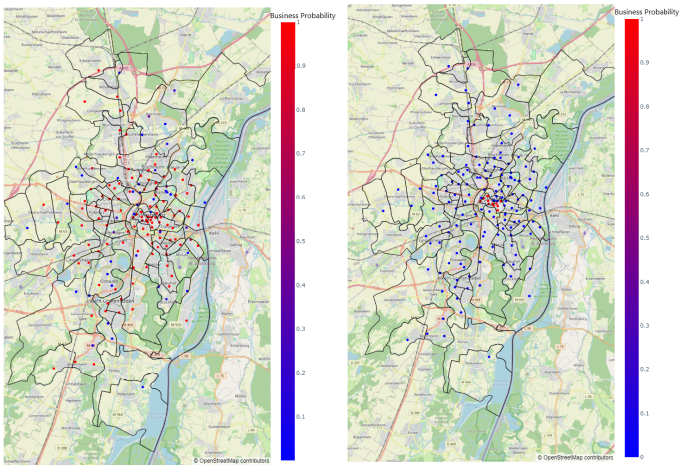

Fig. 3: Comparative maps of Business Traffic Probability in Strasbourg: Tuesday Morning (left) vs. Friday Evening (right)

traffic during the midday period. This pattern likely correlates with tourists' and locals' activities around the cathedral area, where sightseeing and leisure are predominant.

General observations across all cells for Monday highlight that the traffic from midnight to 09:00 is predominantly pleasure-oriented, suggesting activities unrelated to business, such as late-night entertainment or early-morning personal errands. From 09:00 to 17:00, the traffic is decisively business-oriented, corresponding with regular business operations. Post 17:00, the traffic composition becomes area-specific, influenced by the local socio-economic fabric. On Sundays, the pattern shifts, with pleasure-oriented traffic dominating the early hours. The period from 09:00 to 12:00 presents a mixed scenario especially due to businesses that open before noon on Sundays, hinting at a transitional phase where individuals engage in various activities. From 12:00 onwards, the pleasure orientation is pronounced, extending well into the night, reflecting the leisurely pace of the weekend. These patterns not only illustrate the temporal and spatial variability of urban mobility but also emphasize the intricate relationship between urban planning, economic activity, and societal habits. The use of GMM in classifying these patterns showcases their efficacy in capturing the nuanced interplay of factors that define urban traffic trends, offering valuable insights for city planners and policymakers to optimize resource allocation and enhance urban living standards.

C. Heatmap representation and traffic overview

The two maps shown in Figure 3 provide a stark visual contrast between the business traffic probabilities in Strasbourg on a busy Tuesday morning and a lively Friday evening. The left image, captured during the hours of 09:00 to 12:00 on Tuesday, indicates a high concentration of business-related activities. This intensity in business traffic is expected as the workweek is in full swing, with professional engagements and commercial transactions peaking. The bustling downtown areas and business districts are pulsating with commerce and workday routines, showcasing the weekday economic vigor of Strasbourg.

In contrast, the right image illustrates the city's transformation on Friday evening, from 17:00 to midnight. Here, we observe a significant shift in traffic dynamics, as indicated by the color blue spreading across most areas. This change suggests a considerable decrease in business activity, transitioning into a period of leisure and relaxation that heralds the weekend. The urban sprawl that was once a hotspot for corporate hustle now simmers down, with only a few spots of red where late-night work or business events continue. The city's nightlife areas begin to light up, as reflected in the slightly warmer tones in certain zones, revealing the pleasure-oriented side of Strasbourg coming to life.

This comparative analysis not only highlights the daily rhythms of urban life but also underscores the versatility of our classification approach. It is important to note that these maps can also be generated for forecasted data, allowing for anticipatory insights into traffic patterns. Such forecasting capabilities empower city planners and businesses with forward-looking strategies to optimize services and infrastructure in line with predicted traffic trends, ensuring a dynamic and responsive urban environment.

D. Discussion

The results of this study underscore the capability of GMM in dissecting the intricate web of network traffic, presenting a binary classification that is also adaptable across temporal and spatial variations. Our exploration into the realms of business and pleasure-oriented traffic offers a nuanced lens through which the dynamics of network usage can be understood and anticipated. This classification, while simplistic, serves as a stepping stone towards developing a more granular understanding of network demands.

The convergence between the classifications from k-means and GMM further validates the robustness of our approach. The 72.2% congruence between the two methods validates GMM's capabilities in capturing the subtleties of network traffic. Moreover, the adaptability of our classification system across different days and zones emphasizes its potential applicability in a wide range of network scenarios. Looking ahead, the application of this classification system extends beyond mere traffic analysis. In network management, it paves the way for more intelligent and dynamic network slicing.

For instance, traffic deemed as 'business' could be allocated to a slice with higher priority, ensuring uninterrupted and low-latency service for critical applications. Conversely, 'pleasure' traffic might be directed to a slice where bandwidth is managed more flexibly, optimizing network resources without compromising overall user satisfaction. In the realm of cybersecurity, our classification model can serve as a sentinel for anomaly detection. By establishing a normative pattern of traffic within business and pleasure venues, any deviation from these patterns can be flagged for further investigation. For urban planning, our classification facilitates a deeper understanding of the city's pulse by highlighting the ebbs

and flows of digital traffic in correlation with urban activity. By distinguishing between traffic generated for business or pleasure, urban planners can identify which areas of the city are most active at different times of the day or year, and plan infrastructure developments accordingly. By predicting network traffic type, we can ensure optimal resource allocation and enhance the overall user experience. Future work will delve into the integration of these classifications with advanced predictive models, aiming to anticipate network demands with greater precision and agility. The potential of this research to impact the future of network traffic classification and management is substantial, as we continue to push the boundaries of what is possible with machine learning and data analytics in network environments.

V. Conclusion and perspectives

In conclusion, this paper presents a novel approach to network traffic classification using Gaussian Mixture Models. Our methodology has demonstrated its efficacy in classifying traffic in the metropolitan area of Strasbourg with a focus on temporal and user behavior patterns. Furthermore, in network management, precise traffic classification empowers service providers to implement dynamic network slicing. This segmentation of the network into slices allows for tailored QoS levels, meeting diverse requirements of different traffic types. As we continue to refine our approach to network traffic prediction and classification, there are two primary avenues we plan to explore in future work:

1) **Multi-tiered classification:** Moreover, we aim to implement a multi-tiered classification scheme that goes beyond the binary categorization of business and pleasure. This scheme will incorporate additional layers of traffic importance, enabling more granular prioritization of network resources.
2) **Expansion to Traffic Prediction:** Future work will focus on enhancing the accuracy of network volume predictions in real-time, which is critical for the adaptive optimization of telecommunications infrastructure.

Future research will also delve into improving the computational efficiency of our models, enabling them to be deployed on a wider range of devices, including those with limited processing capabilities.

References

[1] Hassan Alizadeh, Abdolrahman Khoshrou, and Andre Zuquete. Traffic classification and verification using unsupervised learning of gaussian mixture models. In *2015 IEEE international workshop on measurements & networking (M&N)*, pages 1–6. IEEE, 2015.

[2] Hongzhi Chen, Ajay Mehra, Stefano Tasselli, and Stephen P Borgatti. Network dynamics and organizations: A review and research agenda. *Journal of Management*, 48(6):1602–1660, 2022.

[3] Adriano AM De Resende, Pedro HAD De Melo, Jefferson R Souza, Renan G Cattelan, and Rodrigo Sanches Miani. Traffic classification of home network devices using supervised learning. In *ICAART (3)*, pages 114–120, 2022.

[4] Jaleh Farmani and Amirreza Khalil Zadeh. Ai-based self-healing solutions applied to cellular networks: An overview. *arXiv preprint arXiv:2311.02390*, 2023.

[5] Kevin Fauvel, Fuxing Chen, and Dario Rossi. A lightweight, efficient and explainable-by-design convolutional neural network for internet traffic classification. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4013–4023, 2023.

[6] Nour Gritli, Ferhat Khendek, and Maria Toeroe. Extending the network service descriptor to capture user isolation intents for network slices. In *2022 IEEE Future Networks World Forum (FNWF)*, pages 340–346. IEEE, 2022.

[7] Zhiping Jin, Zhibiao Liang, Meirong He, Yao Peng, Hanxiao Xue, and Yu Wang. A federated semi-supervised learning approach for network traffic classification. *International Journal of Network Management*, 33(3):e2222, 2023.

[8] Marios Evangelos Kanakis, Ramin Khalili, and Lin Wang. Machine learning for computer systems and networking: A survey. *ACM Computing Surveys*, 55(4):1–36, 2022.

[9] Josef Koumar, Karel Hynek, and Tomáš Čejka. Network traffic classification based on single flow time series analysis. In *2023 19th International Conference on Network and Service Management (CNSM)*, pages 1–7. IEEE, 2023.

[10] Ming Li and Guikai Liu. A network traffic classification model based on xgboost_rfecv feature extraction. In *Proceedings of the 4th International Conference on Advanced Information Science and System*, pages 1–5, 2022.

[11] Massimiliano Maule, John S Vardakas, Georgios Kormentzas, and Christos Verikoukis. Multi-service single tenant 5g fronthaul resource orchestration framework based on network slicing. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 3417–3422. IEEE, 2022.

[12] Machoke Mwita and Johnson Agbinya. Performance comparison of ensemble learning with supervised algorithms in classifying multi-label network traffic flow machoke mwita1,*, johnson agbinya3, jimmy mbelwa 2, and anael elikana sam 4.

[13] Mohammed Nsaif, Gergely Kovásznai, Mohammed Abboosh, Ali Malik, and Ruairí de Fréin. Ml-based online traffic classification for sdns. In *2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, pages 217–222. IEEE, 2022.

[14] Mohammad Rezaee and Mohammad Hossein Yaghmaee Moghaddam. Sdn-based quality of service networking for wide area measurement system. *IEEE Transactions on Industrial Informatics*, 16(5):3018–3028, 2019.

[15] S Saibharath, Sudeepta Mishra, and Chittaranjan Hota. Quality of service driven resource allocation in network slicing. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5. IEEE, 2020.

[16] Muhammad Saqib, Zakaria Ait Hmitti, Halima Elbiaze, and Roch H Glitho. An accurate & efficient approach for traffic classification inside programmable data plane. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 6331–6336. IEEE, 2022.

[17] Randeep Singh, Abolfazl Mehbodniya, Julian L Webber, Pankaj Dadheech, G Pavithra, Mohammed S Alzaidi, Reynah Akwafo, et al. Analysis of network slicing for management of 5g networks using machine learning techniques. *Wireless Communications and Mobile Computing*, 2022, 2022.

[18] Yipeng Wang, Xiaochun Yun, Yongzheng Zhang, Chen Zhao, and Xin Liu. A multi-scale feature attention approach to network traffic classification and its model explanation. *IEEE Transactions on Network and Service Management*, 19(2):875–889, 2022.

[19] Chenguang Zhang, Run Dong, Xiaobin Zhang, and Weilin Li. A comparative analysis of supervised/unsupervised algorithms in phm application. In *2022 Global Reliability and Prognostics and Health Management (PHM-Yantai)*, pages 1–5. IEEE, 2022.

[20] Jielun Zhang, Fuhao Li, and Feng Ye. Network traffic clustering with qos-awareness. *China Communications*, 19(3):202–214, 2022.