

# Przydział zasobów w elastycznych sieciach optycznych z wykorzystaniem metod uczenia maszynowego

Katarzyna Duszyńska, Paweł Polski, Michał Włosek  
Politechnika Wrocławska  
Wrocław, Polska

**Streszczenie**—Praca opisuje rozwiązanie problemu przydziału zasobów w elastycznych sieciach optycznych. Przewidywane są cztery parametry: najwyższy slot, średni najwyższy slot, suma slotów oraz średnia liczba transceiverów. Analizując problem oraz dane stwierdzono, że dobrym podejściem może być redukcja wymiarowości, usuwając cechy, które nie niosą dużej ilości informacji. W tym celu stworzono algorytm bazujący na drzewach decyzyjnych. Sprawdzono wpływ wspomnianego algorytmu na wynik podstawowych modeli regresyjnych. Kolejnym krokiem było porównanie płytkich modeli regresyjnych z tymi bardziej skomplikowanymi. Finalnie w celu porównania algorytmów wykorzystano testy statystyczne t-studenta.

**Index Terms**—RSA, uczenie maszynowe, sieci neuronowe, regresja, elastyczne sieci optyczne

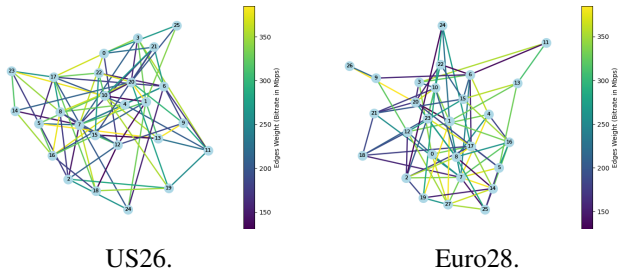
## I. WSTĘP

Od wielu lat obserwowany jest ciągły wzrost ruchu sieciowego. Pandemia Covid-19 przyczyniła się znacząco do wspomnianego wzrostu [3]. W związku z tym zwiększa się również stopień skomplikowania sieci oraz ryzyko wystąpienia wąskiego gardła. Idealnym rozwiązaniem mogą okazać się elastyczne sieci optyczne [9], które umożliwiają przydzielenie pasma długości fali zgodnie z aktualnym zapotrzebowaniem (flex-grid) [17]. Jednak głównym problemem jest optymalizacja ilości pasma potrzebnego do aktualnej transmisji bez blokowania ruchu. Problem ten w literaturze można spotkać jako problem przydziału zasobów i routingu (RSA) [8]. Odpowiednia optymalizacja jest istotna, gdyż pozwala zmniejszyć zużycie energii, oraz ilość użytych transceiverów, a co za tym idzie koszty. Ciężko jednak znaleźć optymalne rozwiązanie dla tego problemu i jest on tematem wielu prac naukowych. Początkowo wykorzystywano do tego celu między innymi algorytmy liniowe które były wystarczające, aby zapewnić odpowiedni poziom otrzymywanych wyników. Jednakże nieliniowy wzrost ilości generowanych przez społeczeństwo danych doprowadził do spadku ich skuteczności. Wraz z popularyzacją uczenia maszynowego zaczęto wykorzystywać popularne metody z tej dziedziny, w celu pośredniego rozwiązania problemu RSA. Modele uczenia maszynowego pozwalają na predykację skutków na podstawie aktualnie posiadanych danych. W celu stworzenia skutecznego modelu potrzebne są duże ilości danych, które pozwolą zgeneralizować problem. Wiele dużych firm jednak niechętnie dzieli się takimi danymi. Istotną kwestią jest czas predykcji, który powinien być jak najmniejszy, lecz wraz

z zastosowaniem bardziej skomplikowanych modeli uczenia maszynowego czas ten może rosnąć. Podsumowując praca skupia się na predykcji czterech parametrów w elastycznych sieciach optycznych, takich jak najwyższy slot, średni najwyższy slot, suma wszystkich slotów, średnia liczba aktywnych transceiverów. Parametry te są wynikiem rozwiązania problemu RSA. Aktualny rozdział jest wprowadzeniem w tematykę, w kolejnym rozdziale analizowane są prace o podobnej tematyce. W rozdziale trzecim omawiany jest problem, który jest podejmowany w tej pracy. Następnie w rozdziale czwartym omawiane są algorytmy, które zostały zaimplementowane oraz autorskie, zaproponowane rozwiązanie.

## II. PRZEGLĄD LITERATURY

Temat optymalizacji sieci optycznych jak i predykcji ruchu sieciowego jest chętnie podejmowany przez naukowców z całego świata. Prace skupiają się między innymi na predykcji ruchu sieciowego [5] [16] [11] [12]. Ma to na celu optymalizować sieć oraz urządzenia w niej pracujące zgodnie z przewidywanym zapotrzebowaniem. Używane są do tego różnego rodzaju modele regresyjne jak i głębokie sieci neuronowe. Często wykorzystuje się specjalistyczne architektury *Long Short-Term Memory*, które wykazują mały błąd predykcji [18]. Kolejnymi obszarami badań są jakość transmisji (*QoT*) [1] i minimalizacja skutków awarii węzła w sieci [2] [10] lub uszkodzeń fizycznych infrastruktury [6]. Oprócz wyżej wymienionych tematów często podejmowana jest tematyka alokacji zasobów [4], który jest tożsamy z problemem podejmowanym w tej pracy. W pracy Shrinivasa Petale i Suresha Subramaniama podejmowany jest temat predykcji parametru  $\gamma^d$  w architekturze Euro28 [14]. Dane wejściowe są różne, lecz idea podobna gdyż w wyżej wymienionym artykule podjęta jest próba predykcji parametru  $\gamma^d$  w celu optymalnej alokacji zasobów. W artykule Yongjuna Zhanga [19] przedstawiono wykorzystanie *ML* w celu przewidywania przesłuchów, szacowania ruchu i *QoT* co pomoże w wyznaczaniu tras i alokacji zasobów sieciowych. Poruszono również pozostałe obszary zastosowań metod uczenia maszynowego w kontekście sieci komputerowych. W artykule [20] przedstawiono algorytm przywracania usług oparty na *A2CSR*. Wykorzystuje on model rozpoznawania obrazu *MobileNetV2* oraz uczenie maszynowe w celu lepszego rozwiązywania sytuacji krytycznych w sieci.



Rysunek 1: Grafy przedstawiające architektury Euro28 i US26 dla jednego wybranego losowo zbioru zapytań.

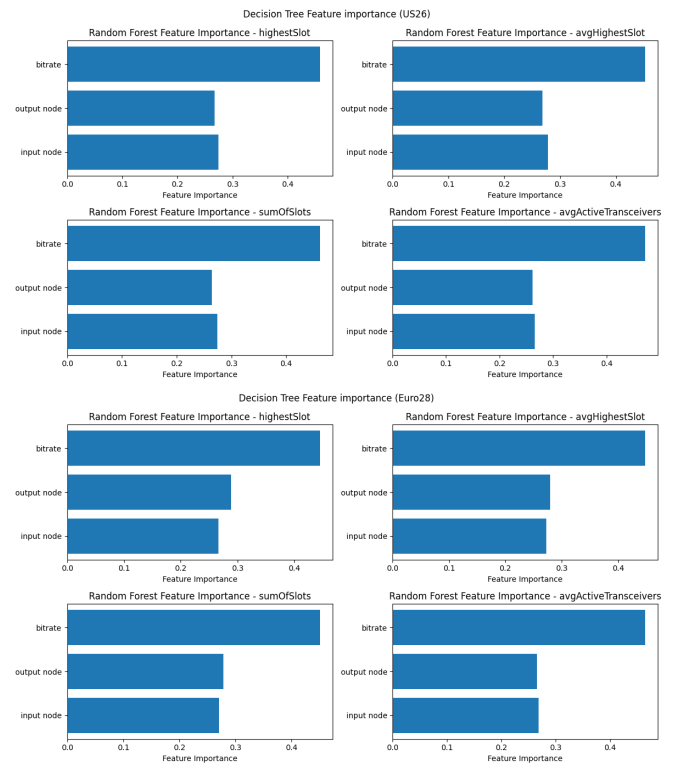
Algorytmy *ML* znajdują również swoje wykorzystanie w dziedzinie automatyzacji sieci [15], gdzie pozwalają na skrócenie procesów konfiguracji i zmniejszenie ryzyka popełnienia błędów przez administratora. Problematyka podejmowana w ramach tej pracy różni się od tej poruszanej w literaturze badanym parametrem.

### III. DEFINICJA PROBLEMU

Problem prezentowany w pracy dotyczy predykcji czterech parametrów, które umożliwiają rozwiązanie problemu przydziału zasobów w elastycznych sieciach optycznych. Parametry te to następujące: najwyższy slot, średni najwyższy slot, suma wszystkich slotów oraz średnia liczba aktywnych transceiverów. Do eksperymentów wybrano dwa znane z literatury zbiory danych: Euro28 i US26 [7]. Reprezentują one 100 zbiorów zapytań, w skład pojedynczego zbioru wchodzi 100 pojedynczych zapytań składających się z węzła wejściowego oraz wyjściowego i przepływności pomiędzy węzłami. Na rysunku 1 przedstawiono wizualną reprezentację danych dla jednego zbioru stu zapytań w postaci grafu dla obydwu zbiorów danych.

Pierwszym napotkanym problemem było znalezienie odpowiedzi na pytanie czy wszystkie dane występujące w zbiorze są istotne w kontekście modelu oraz jak je przekształcić aby nadawały się do uczenia. Następnie podjęto próbę oszacowania istotności każdego z trzech parametrów wejściowych w kontekście każdego przewidywanego parametru. Przeprowadzono proste badania, w których dla każdego zbioru zapytań określono istotność każdego z parametrów. Finalnie uśredniono wyniki aby pokazać statystykę dla całego zbioru.

Z wykresów przedstawionych na rysunku numer 2 można odczytać, że najwięcej informacji niesie parametr jakim jest przepływność. Pozostałe parametry przedstawiają podobny poziom istotności, lecz jest on znacznie niższy niż w pierwszym przypadku. Podsumowując, po zbadaniu informacji, niesionej w każdym parametrze uznano, że wszystkie 3 parametry są istotne w kontekście uczenia modelu. Dlatego surowe, wczytane dane przekształcono z rozmiaru tensora  $100 \times 100 \times 3$  do rozmiaru tensora  $100 \times 300$ . Tworzy to dużą przestrzeń cech i pomocne może okazać się wykorzystanie metody selekcji cech w przypadku użycia modeli płytkich. Modele głębokie posiadają już same w sobie taką selekcję więc w tym przypadku wydaje się to zbędne.



Rysunek 2: Wykresy przedstawiają istotność każdego z 3 parametrów wejściowych.

### IV. ALGORYTMY

Po dogłębnej analizie literatury stwierdzono, że w zastosowaniach predykcji przydziału zasobów elastycznych sieci optycznych bardzo często stosuje się różne odmiany drzew decyzyjnych, algorytm k najbliższych sąsiadów oraz maszyny wektorów nośnych. Dodatkowo coraz częściej spotyka się głębokie sieci neuronowe w tych zastosowaniach. W ramach tej pracy podjęto dwa tematy. Pierwszy to temat zbadania wpływu wyboru cech w kontekście prostych/płytkich modeli regresyjnych. W tym celu zaproponowano wykorzystanie wiedzy pochodzącej z drzewa decyzyjnego wyuczonego na analizowanych danych w celu uzyskania tych cech, które zostaną uznane za najbardziej istotne. Algorytm stworzony w tym celu można przedstawić krokami z algorytmu 1.

W kolejnym kroku została podjęta próba wykorzystania sieci neuronowych w celu predykcji szukanych parametrów. Dostępne dane można przedstawić jako dane tabelaryczne, więc wykorzystanie konwolucyjnych sieci neuronowych wydaje się zasadne. Bardzo popularne w zastosowaniach predykcji ruchu sieciowego są również rekurencyjne sieci neuronowe dlatego wykorzystano również sieci neuronowe *LSTM* i *GRU*. W tabeli I przedstawiono wykorzystane parametry w użytych algorytmach. Jeżeli jakiegoś parametru nie podano w tabeli to oznacza to, że przyjęto domyślną wartość. W tych modelach wykorzystano błąd średnio kwadratowy jako funkcje strat, jako metrykę wykorzystano średnie bezwzględne odchylenie procentowe oraz użyto algorytmu *Adam* jako optymalizatora.

**Algorithm 1** *FeatureSelection***Input:** $D$  - Dane przedstawiające ruch sieciowy**Symbols:** $FI$  - Istotność cech uzyskana z algorytmu drzew decyzyjnych $IC$  - Tablica indeksów cech, które niosą informacje $c$  - Cechy pochodzące z zbioru danych**Output:** $D_t$  - Dane po transformacji z zmniejszoną liczbą cech

```

function FEATURESELECTION( $D$ )
   $FI \leftarrow DecisionTreeRegressor(D)$ 
   $IC \leftarrow \emptyset$ 
  for  $c \leftarrow D$  do
    if  $FI[c] > 0$  then
       $IC += c$ 
    end if
  end for
   $return \leftarrow D_t = D[IC]$ 
end function

```

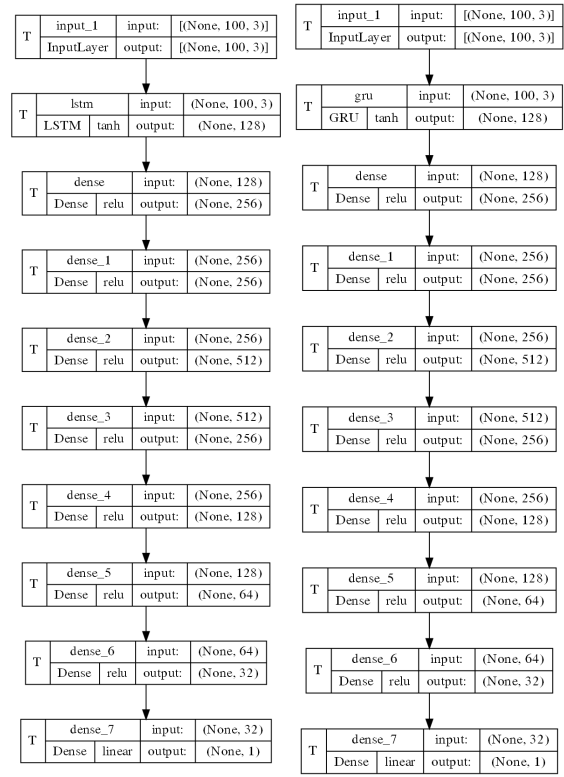
Stworzone w ramach pracy modele i ich architektury przedstawiono na rysunkach 4 i 3.

Regresor	Wykorzystane parametry
CART	random_state=1410
SVR	kernel=poly
KNN	k_neighbours=5
RF	random_state=1410
MLP	hidden_layer_sizes=50, batch_size=25, random_state=1410, warm_start=True
LR	-
LSTM	learning_rate=0.001, loss="mse"
GRU	learning_rate=0.001, loss="mse"
CNN	learning_rate=0.001, loss="mse"

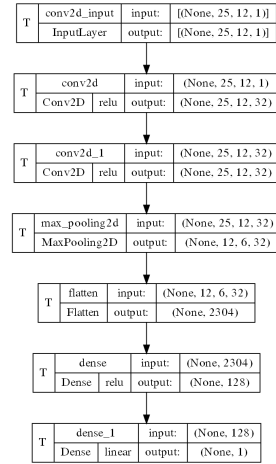
Tabela I: Wykorzystane parametry regresorów podczas eksperymentu.

**V. PROTOKÓŁ EKSPERYMENTACYJNY**

Wybrane w poprzednim etapie metody poddaliśmy pod eksperyment, którego protokół rozpoczął się od  $k$ -foldowej walidacji krzyżowej. Metoda ta dzieli wykorzystywany zbiór na  $k$  części, wykorzystując jedną z nich do procesu testowania, a pozostałe podzbiory są przeznaczane do trenowania modelu. Proces jest powtarzany do momentu wykorzystania wszystkich foldów do celów uczenia. Pozwala ona na uzyskanie lepszej jakości oceny wydajności niż w przypadku zastosowania pojedynczego podziału. Podczas tego eksperymentu ustalono wartość  $k$  równą 5. Eksperyment został podzielony na dwie części. W ramach jednej z nich zastosowano selekcję cech opisaną algorytmem nr 1. W następnych etapach badane modele zostały wyuczone na danych treningowych, po czym dokonano predykcji wartości badanych rozwiązań.



Rysunek 3: Architektura wykorzystanych sieci LSTM i GRU.



Rysunek 4: Architektura wykorzystanej sieci CNN.

W celu oceny rezultatów przeprowadzonych badań wykorzystano metrykę *MAPE* - *mean absolute percentage error* (średnie bezwzględne odchylenie procentowe). Informuje ona o średniej wielkości błędów prognoz, pozwalając na porównanie różnych metod.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \quad (1)$$

gdzie  $A_t$  to rzeczywista wartość w chwili  $t$ ,  $F_t$  to wartość prognozowana w chwili  $t$ , a liczba obserwacji to  $n$ . Do

przeprowadzenia eksperymentu wykorzystano funkcję z biblioteki sklearn [13], dlatego w porównaniu do wzoru nr 1 nie jest on podawany jako wartość procentowa. W kolejnym kroku przeprowadzone zostały testy statystyczne t-studenta, pozwalające stwierdzić czy różnice pomiędzy otrzymanymi rezultatami są istotne statystycznie. Wynik uznano za istotny statystycznie, gdy poziom istotności jest mniejszy niż 0,05.

## VI. WYNIKI

W tabelach II, IV, VI oraz VII przedstawiono wyniki testu t-studenta dla metryki *MAPE1* predykowanych parametrów. Kolorem żółtym zostały zaznaczone algorytmy, które statystycznie najlepiej poradziły sobie z zadaniem problemem. Parametr określający średnią liczbę aktywnych transceiverów dla każdego przypadku osiąga najmniejszy błąd, co oznacza że teoretycznie najłatwiej go przewidzieć. W eksperymencie 1 dla obu zbiorów tylko dla jednego parametru udało się określić statystycznie najlepszy regresor. Zarówno w zbiorze euro28 i us26 najlepszym statystycznie regresorem dla parametru średnia liczba aktywnych transceiverów okazał się być *SVR* (*Support Vector Regression*).

MAPE										
Parametr	CART	SVR	KNN	RF	LR	CART-FS	SVR-FS	KNN-FS	RF-FS	LR-FS
highestSlot	0.125	0.094	0.093	0.092	0.146	0.131	0.091	0.094	0.099	0.109
	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 5	—
avgHighestSlot	0.104	0.067	0.074	0.074	0.100	0.105	0.071	0.077	0.075	0.186
	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
sumOfSlots	0.072	0.051	0.054	0.053	0.068	0.072	0.051	0.057	0.055	0.117
	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
avgActiveTransceivers	0.032	0.020	0.025	0.023	0.031	0.032	0.023	0.026	0.024	0.031
	—	1, 3, 5, 6, 8, 9, 10	1, 5, 6, 10	1, 5, 6, 10	—	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—

Tabela II: Wyniki testu t-studenta dla metryki *MAPE* predykowanych parametrów w zbiorze euro28 - eksperyment 1.

Czas wykonania					
Parametr	CART	KNN	SVR	RF	LR
highestSlot	0.020	0.004	0.020	1.181	0.013
avgHighestSlot	0.025	0.004	0.015	1.263	0.010
sumOfSlots	0.022	0.004	0.005	1.667	0.009
avgActiveTransceivers	0.026	0.004	0.010	1.442	0.023

Czas wykonania					
Parametr	CART-FS	KNN-FS	SVR-FS	RF-FS	LR-FS
highestSlot	0.004	0.004	0.020	0.383	0.006
avgHighestSlot	0.006	0.003	0.003	0.635	0.005
sumOfSlots	0.007	0.003	0.006	0.580	0.005
avgActiveTransceivers	0.003	0.002	0.006	0.493	0.003

Tabela III: Czas działania algorytmów dla zbioru euro28 - eksperyment 1.

t-studenta										
Parametr	CART	SVR	KNN	RF	LR	CART-FS	SVR-FS	KNN-FS	RF-FS	LR-FS
highestSlot	0.197	0.132	0.145	0.146	0.249	0.192	0.135	0.150	0.147	0.207
	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 5	—
avgHighestSlot	0.133	0.085	0.090	0.089	0.130	0.133	0.086	0.094	0.093	0.270
	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
sumOfSlots	0.074	0.054	0.054	0.055	0.084	0.075	0.055	0.058	0.057	0.119
	5, 10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
avgActiveTransceivers	0.033	0.022	0.023	0.023	0.036	0.032	0.024	0.027	0.024	0.035
	5	1, 5, 6, 7, 8, 10	1, 5, 6, 8, 10	1, 5, 6, 8, 10	—	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—

Tabela IV: Wyniki testu t-studenta dla metryki *t-studenta* predykowanych parametrów w zbiorze us26 - eksperyment 1.

Czas wykonania					
Parametr	CART	KNN	SVR	RF	LR
highestSlot	0.025	0.003	0.007	1.720	0.039
avgHighestSlot	0.025	0.004	0.018	1.792	0.026
sumOfSlots	0.023	0.003	0.003	1.397	0.009
avgActiveTransceivers	0.022	0.009	0.007	1.347	0.015

Czas wykonania					
Parametr	CART-FS	KNN-FS	SVR-FS	RF-FS	LR-FS
highestSlot	0.005	0.002	0.004	0.463	0.003
avgHighestSlot	0.007	0.002	0.003	0.521	0.004
sumOfSlots	0.006	0.002	0.005	0.547	0.015
avgActiveTransceivers	0.004	0.002	0.004	0.398	0.002

Tabela V: Czas działania algorytmów dla zbioru us26 - eksperyment 1.

W eksperymencie 2 dla obu zbiorów udało się określić statystycznie najlepszy regresor dla dwóch parametrów: suma slotów oraz średnia liczba aktywnych transceiverów. Zarówno w zbiorze euro28 i us26 najlepszym statystycznie regresorem dla obu parametrów okazał się być *CNN* (*Convolutional Neural Network*).

MAPE									
Parametr	CART	SVR	KNN	RF	LR	MLP	LSTM	GRU	CNN
highestSlot	0.114	0.091	0.095	0.093	0.113	0.436	0.042	0.038	0.022
	6	1, 5, 6	1, 6	1, 5, 6	6	—	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6
avgHighestSlot	0.101	0.071	0.077	0.075	0.177	0.611	0.019	0.022	0.018
	5, 6	1, 5, 6	1, 5, 6	1, 5, 6	6	—	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6
sumOfSlots	0.071	0.051	0.053	0.051	0.112	0.557	0.024	0.036	0.011
	5, 6	1, 5, 6	1, 5, 6	1, 5, 6	6	—	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, 8
avgActiveTransceivers	0.033	0.023	0.025	0.024	0.03	5.568	0.021	0.020	0.011
	6	1, 5, 6	1, 6	1, 5, 6	6	—	1, 5, 6	1, 5, 6	all

Tabela VI: Wyniki testu t-studenta dla metryki *MAPE* predykowanych parametrów w zbiorze euro28 - eksperyment 2.

MAPE									
Parametr	CART	SVR	KNN	RF	LR	MLP	LSTM	GRU	CNN
highestSlot	0.190	0.137	0.155	0.146	0.201	0.241	0.063	0.053	0.033
	—	1, 5	5	1, 5	—	—	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6
avgHighestSlot	0.136	0.089	0.098	0.094	0.230	0.506	0.040	0.033	0.020
	5, 6	1, 5, 6	1, 5, 6	1, 5, 6	6	—	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6
sumOfSlots	0.076	0.055	0.059	0.058	0.129	0.545	0.028	0.043	0.010
	5, 6	1, 5, 6	1, 5, 6	1, 5, 6	6	—	1, 2, 3, 4, 5, 6, 8	1, 2, 3, 4, 5, 6	all
avgActiveTransceivers	0.032	0.025	0.025	0.025	0.036	4.979	0.026	0.025	0.006
	6	1, 5, 6	1, 5, 6	1, 5, 6	6	—	5, 6	1, 5, 6	all

Tabela VII: Wyniki testu t-studenta dla metryki *MAPE* predykowanych parametrów w zbiorze us26 - eksperyment 2.

## VII. WNIOSKI

W pierwszym eksperymencie podjęto próbę oceny, zasadności stosowania przetwarzania wstępnego w kontekście prostych modeli. Rozważano przetwarzanie wstępne tylko w kontekście prostych modeli, gdyż w modelach bardziej skomplikowanych jakimi są sieci neuronowe, ekstrakcja cech jest wbudowana w proces uczenia. Z otrzymanych wyników wynika, że wpływ selekcji cech na wyniki regresji nie jest istotny statystycznie. Mimo wszystko fakt, że przy wykorzystaniu mniejszej ilości cech uzyskano podobne wyniki, pozwala na uzyskanie porównywalnych rezultatów przy zastosowaniu krótszego procesu obliczeniowego. Wykorzystanie mniejszej ilości cech skutecznie oszczędza czas uczenia oraz predykcji wykorzystanych modeli. Zmniejszenie czasu wykonania oznacza również zmniejszenie złożoności obliczeniowej co jest

korzystne w kontekście sieci komputerowych i idei jaką jest green networking. Mówi ona o minimalizacji emisji dwutlenku węgla w kontekście sieci komputerowych. Oznacza to, że im większa złożoność algorytmu działającego w sieci, tym większe zapotrzebowanie na energię, a co za tym idzie większa emisja dwutlenku węgla.

Drugim eksperymentem, który przeprowadzono jest porównanie wielu modeli (prostych jak i tych głębokich) ze sobą. Z wyników testów statystycznych można odczytać, że z tym zadaniem znacznie lepiej radzą sobie modele głębokie. Najlepiej statystycznie radzi sobie model oparty o konwolucyjne sieci neuronowe.

Analizując wszystkie wykonane eksperymenty można stwierdzić, że mniejsze wyniki błędu *MAPE* otrzymywane są na zbiorze euro28, jednakże nie można tego potwierdzić gdyż nie wykonano testów statystycznych w tym zakresie. Dodatkowo zauważono że wiedza zawarta w danych uczących najlepiej opisuje parametr średniej liczby transceiverów.

Podsumowując stwierdzono, że najlepiej z przedstawionym zadaniem poradziły sobie głębokie modele, jednak ich złożoność obliczeniowa oraz czas działania często mogą być wykluczające w specyficznych zastosowaniach w sieciach komputerowych. Modele płytkie z wykorzystanym mechanizmem selekcji cech mogą okazać się dobrym wyborem w przypadkach gdy zależy nam na lekkim modelu o stosunkowo szybkim działaniu.

## LITERATURA

- [1] Luca Barletta, Alessandro Giusti, Cristina Rottondi, and Massimo Tornatore. Qot estimation for unestablished lighpaths using machine learning. In *Optical Fiber Communication Conference*, page Th1J.1. Optica Publishing Group, 2017.
- [2] Xiaoliang Chen, Che-Yu Liu, Roberto Proietti, Zhaohui Li, and S. J. Ben Yoo. Automating optical network fault management with machine learning. *IEEE Communications Magazine*, 60(12):88–94, 2022.
- [3] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poesse, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, et al. The lockdown effect: Implications of the covid-19 pandemic on internet traffic. In *Proceedings of the ACM internet measurement conference*, pages 1–18, 2020.
- [4] Rahul Jashvantbhai Pandya. Machine learning-oriented resource allocation in c+ l+ s bands extended sdm-eons. *IET Communications*, 14(12):1957–1967, 2020.
- [5] Manish Joshi and Theyazn Hassn Hadi. A review of network traffic analysis and prediction techniques. *arXiv preprint arXiv:1507.05722*, 2015.
- [6] Florian Azendorf Helmut Griesser Carsten Tropschug Khoulood Abdelli, Joo Yeon Cho and Stephan Pachnicke. Machine-learning-based anomaly detection in optical fiber monitoring. *IET Communications*, 14(5):365–375, 2022.
- [7] Mirosław Klinkowski and Krzysztof Walkowiak. On the advantages of elastic optical networks for provisioning of cloud computing traffic. *IEEE Network*, 27(6):44–51, 2013.
- [8] Mirosław Klinkowski, Mateusz Zotkiewicz, Krzysztof Walkowiak, Michał Pióro, Marc Ruiz, and Luis Velasco. Solving large instances of the rsa problem in flexgrid elastic optical networks. *Journal of Optical Communications and Networking*, 8(5):320–330, 2016.
- [9] Victor López, Luis Velasco, et al. Elastic optical networks. *Architectures, Technologies, and Control, Switzerland: Springer Int. Publishing*, 2016.
- [10] Francesco Musumeci. Machine learning for failure management in optical networks. In *Optical Fiber Communication Conference*, pages Th4J–1. Optica Publishing Group, 2021.
- [11] Francesco Musumeci, Cristina Rottondi, Avishek Nag, Irene Macaluso, Darko Zibar, Marco Ruffini, and Massimo Tornatore. An overview on application of machine learning techniques in optical networks. *IEEE Communications Surveys & Tutorials*, 21(2):1383–1408, 2018.
- [12] Tiago Prado Oliveira, Jamil Salem Barbar, and Alessandro Santos Soares. Computer network traffic prediction: a comparison between traditional and deep learning neural networks. *International Journal of Big Data Intelligence*, 3(1):28–37, 2016.
- [13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [14] Shrinivas Petale and Suresh Subramaniam. Machine learning-aided resource assignment in space division multiplexed elastic optical networks. In *Proc. International Conference on Computer Communications and Networks (ICCCN)*, 2021.
- [15] Danish Rafique and Luis Velasco. Machine learning for network automation: Overview, architecture, and applications. *Journal of Optical Communications and Networking*, 10(5):D126–D143, 2018.
- [16] Nipun Ramakrishnan and Tarun Soni. Network traffic prediction using recurrent neural networks. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 187–193. IEEE, 2018.
- [17] L. Velasco, M Klinkowski, M Ruiz, V López, and G Junyent. Elastic spectrum allocation for variable traffic in flexible-grid optical networks. In *Optical Fiber Communication Conference*, pages JTh2A–39. Optica Publishing Group, 2012.
- [18] Shihao Wang, Qinzhen Zhuo, Han Yan, Qianmu Li, and Yong Qi. A network traffic prediction method based on lstm. *ZTE Communications*, 17(2):19–25, 2019.
- [19] Yongjun Zhang, Jingjie Xin, Xin Li, and Shanguo Huang. Overview on routing and resource allocation based machine learning in optical networks. *Optical Fiber Technology*, 60:102355, 2020.
- [20] Yajie Li Feng Wang Xinghua Li Dahai Han Zipiao Zhao, Yongli Zhao and Jie Zhang. Service restoration in multi-modal optical transport networks with reinforcement learning. *Journal of Optical Communications and Networking*, 29:3825–3840, 2021.