



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Διπλωματική Εργασία

Εφαρμογή Τεχνικών Μηχανικής Μάθησης για Ταξινόμηση Χρηστών σε Δίκτυα 5^{ης} Γενιάς

Μάριος Χαραλαμπίδης
Α.Μ. 1056925

Επιβλέπων
Χρήστος Μπούρας

Συνεπιβλέπων
<Όνοματεπώνυμο>, <Τίτλος>

Μέλοι Επιτροπής Αξιολόγησης
<Όνοματεπώνυμο>, <Τίτλος>

Πάτρα, 2022

© Copyright συγγραφέας Μάριος Χαραλαμπίδης, 2022

© Copyright θέματος Μάριος Χαραλαμπίδης

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Πατρών δεν υποδηλώνει απαραίτητα και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Χρήστο Ι. Μπούρα και τον Δρ. Απόστολο Γκάμα για την επίβλεψη αυτής της διπλωματικής εργασίας. Είμαι ιδιαίτερα ευγνώμων για την καθοδήγηση και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους συγγενείς και φίλους μου που βρισκόντουσαν στο πλευρό μου όλα αυτά τα χρόνια και στις καλές και στις δύσκολες στιγμές, και με καθοδήγησαν στο να μπορώ να αποτυπώσω τις ιδέες μου όσο πιο επιστημονικώς ορθά γίνεται.

Πρόλογος

Η παρούσα εργασία αποτελεί διπλωματική εργασία στα πλαίσια της προπτυχιακής φοίτησης στο τμήμα των Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής (ΤΜΗΥΠ) του Πανεπιστημίου Πατρών. Ο τίτλος της διπλωματικής αυτής εργασίας είναι «Εφαρμογή Τεχνικών Μηχανικής Μάθησης για ταξινόμηση χρηστών σε Δίκτυα 5ης Γενιάς».

Περίληψη

Τα Κινητά Δίκτυα 5ης Γενιάς αποτελούν την τελευταία τεχνολογία στον κόσμο των Δικτύων Κινητής Τηλεφωνίας. Έχουν την δυνατότητα να προσφέρουν αυξανόμενες ταχύτητες, σταθερές και αξιόπιστες συνδέσεις, ελάχιστη καθυστέρηση και περισσότερη κάλυψη ως προς το πλήθος των συνδεδεμένων συσκευών. Αυτή η αυξανόμενη κάλυψη έχει ως αποτέλεσμα οι Σταθμοί Βάσης να γίνονται όλο και πιο πυκνοί. Το 5G πλέον είναι σε θέση να προσφέρει νέες τεχνολογίες όπως Multiple-Input και Multiple-Output (MIMO), Downlink και Uplink Decoupling (DUDe) και ενσωμάτωση Μηχανικής Μάθησης (ML). Αυτές οι τεχνολογίες έχουν συνεισφέρει σε μεγάλο βαθμό στην βελτιστοποίηση θεμάτων όπως την Ταξινόμηση και Ανάθεση Χρηστών σε Δίκτυα 5G αλλά και την διαχείριση των πόρων εντός τους δικτύου. Αυτό που θα επιχειρήσουμε να δείξουμε είναι πως διάφορες Τεχνικές Μηχανικής Μάθησης όπως Αλγόριθμοι Συσταδοποίησης και Αλγόριθμοι Παλινδρόμησης μπορούν να συνεργαστούν για την καλύτερη ανάθεση των χρηστών, οδηγώντας σε μεγαλύτερους ρυθμούς αποστολής δεδομένων και μεγαλύτερη κάλυψη συσκευών. Η εφαρμογή έχει γίνει σε ένα κινητό δίκτυο επικοινωνίας που έχει δειγματοληπτηθεί από δεδομένα πραγματικών χρηστών. Για τους σκοπούς αυτής της διπλωματικής έχει αναπτυχθεί ένα περιβάλλον προσομοίωσης που αναπαριστά Συσκευές Χρηστών και Σταθμούς βάσης σε πραγματικές συντεταγμένες κάνοντας εφικτή την μελέτη μετρικών ποιότητας κινητών δικτύων σε σενάρια πραγματικού περιβάλλοντος.

Abstract

5th Generation Mobile Networks are the latest technology in the world of Mobile Telephony Networks. They have the ability to offer increased speeds, stable and reliable connections, minimal latency and more coverage on the number of connected devices. This increasing coverage has resulted in Base Stations becoming more and more dense. 5G is now able to offer new technologies such as Multiple-Input and Multiple-Output (MIMO), Downlink and Uplink Decoupling and Machine Learning (ML) utilization. These technologies have greatly contributed to the optimization of issues such as Classification and Assignment of Users to 5G Networks as well as the management of resources within their network. What we will try to demonstrate here is that various Machine Learning Techniques such as Clustering Algorithms and Regression Algorithms can work side by side for the better user engagement leading to higher data rates and greater coverage of devices on a mobile communication network sampled from real user data. For the purposes of this dissertation, a simulation environment has been developed that represents User Equipment and Base Stations in real geolocation coordinates and estimated metrics of mobile networks.

Περιεχόμενα

Κεφάλαιο 1. Εισαγωγή.....	1
1.1 Προηγούμενες Γενιές Κινητών Δικτύων Επικοινωνίας	1
1.2 Δίκτυα 5ης Γενιάς.....	3
1.2.1 Αρχιτεκτονική Δικτύων 5ης Γενιάς	3
1.2.2 Τεχνολογίες Δικτύων 5ης Γενιάς	5
Κεφάλαιο 2. Μηχανική Μάθηση.....	11
2.1 Τεχνικές Μηχανικής Μάθησης	11
2.2 Μάθηση με Επίβλεψη	13
2.2.1 Κατηγοριοποίηση	13
2.2.2 Παλινδρόμηση	14
2.3 Μάθηση Χωρίς Επίβλεψη.....	16
2.3.1 Συσταδοποίηση	16
2.4 Ενισχυτική Μάθηση	24
2.5 Βαθιά Μάθηση (Deep Learning)	25
Κεφάλαιο 3. Περιγραφή Γραφικού Περιβάλλοντος.....	27
3.1 Δημιουργία Γραφικού Περιβάλλοντος.....	27
3.2 Συσταδοποίηση Χρηστών	42
3.3 Συσχετισμός Χρηστών με Σταθμούς Βάσης.....	44
Κεφάλαιο 4. Παρουσίαση Αποτελεσμάτων.....	49
4.1 Σύνοψη.....	49
4.1 Συσταδοποίηση Χρηστών	49
4.2.1 Elbow Method	49
4.2.2 Silhouette Method.....	52
4.2.3 Συσχετισμός UE με BS.....	58
4.2.4 Μελέτη ως προς την Βελτιστοποίηση του Ρυθμού Δεδομένων	61
Κεφάλαιο 5. Συμπεράσματα και Μελλοντικό Έργο	64
5.1 Σύνοψη.....	64
5.2 Μελλοντική Εργασία.....	63
Βιβλιογραφία- Αναφορές	65
Σύντομο Βιογραφικό Συγγραφέα.....	68

Λίστα Εικονών

Εικόνα 1 Συγκριτικός πίνακας προηγούμενων γενιών	2
Εικόνα 2. Resource Grid	4
Εικόνα 3. Network Slicing	7
Εικόνα 4. Η Αρχιτεκτονική ενός SDN Συστήματος.....	8
Εικόνα 5. Massive MIMO Network	9
Εικόνα 6. Κατηγορίες Μηχανικής Μάθησης.....	12
Εικόνα 7. Αλγόριθμοι Κατηγοριοποίησης	13
Εικόνα 8. Linear Regression	14
Εικόνα 9. Random Forest	15
Εικόνα 10. Αλγόριθμος K-means	17
Εικόνα 11. Παραγόμενες Συστάδες.....	18
Εικόνα 12. CF-Tree	23
Εικόνα 13 Βρόχος Ανταμοιβής	24
Εικόνα 14. Νευρωνικό Δίκτυο	25
Εικόνα 15. NN LSTM	26
Εικόνα 16. Γραφικό Περιβάλλον	27
Εικόνα 17. Απεικόνιση σημείων στον Χάρτη	32
Εικόνα 18 Πίνακας Στοιχείων BS	35
Εικόνα 19. Start Screen of the Simulation.....	44
Εικόνα 20. Διάταξη BSs με Kmeans για n=50, n=800.....	60
Εικόνα 21. Διάταξη BSs με PAM για n=50, n=800	60
Εικόνα 22. Διάταξη BSs με Birch για n=50, n=800.....	60

Λιστα Σχημάτων

Σχήμα 1.Μηχανισμός Handover.....	48
Σχήμα 2. Elbow Method K-means	50
Σχήμα 3.Elbow Method K-medoids (PAM).....	51
Σχήμα 4.Elbow Methods Birch	52
Σχήμα 5. K-means, users=50, Average silhouette_score is : 0.44138870374298655	53
Σχήμα 6. K-means, users=100, Average silhouette_score is : 0.43380580197503243	53
Σχήμα 7. K-means, users=200, Average silhouette_score is : 0.3771023809015711	54
Σχήμα 8. K-means, users=400, Average silhouette_score is : 0.37622084556785224	54
Σχήμα 9. K-means, users=800, Average silhouette_score is : 0.35068334190816214	54
Σχήμα 10. PAM, users=50, The average silhouette_score is : 0.36619911250743536	55
Σχήμα 11. PAM, users=100, Average silhouette_score is : 0.3652894167234125	55
Σχήμα 12. PAM, users=200, Average silhouette_score is : 0.3872290003478699	55
Σχήμα 13. PAM, users=400, Average silhouette_score is : 0.3494148478152864	56
Σχήμα 14. PAM, users=800, Average silhouette_score is : 0.37031077201504997	56
Σχήμα 15. Birch, users=50, Average silhouette_score is : 0.422720559467441	56
Σχήμα 16. Birch, users=100, Average silhouette_score is : 0.32339607482356636	57
Σχήμα 17. Birch, users=200, Average silhouette_score is : 0.3519887759110538	57
Σχήμα 18. Birch, users=400, Average silhouette_score is : 0.36518402227632424	57
Σχήμα 19. Birch, users=800, Average silhouette_score is : 0.3206565291438612	58
Σχήμα 20. Ποσοστά Κάλυψης UE	59
Σχήμα 21. Ποσοστά Κάλυψης UE	60
Σχήμα 22. Μέσο Throughput ανα BS.....	62
Σχήμα 23. Μέσο Throughput με Handover	63

Λίστα Πινάκων

Πίνακας 1. CQI and SINR mapping.....	37
Πίνακας 2. MCS index Table	38
Πίνακας 3. 5G NR Scaling Factor	39
Πίνακας 4. 5G NR Numerology	39
Πίνακας 5. Παραμέτροι Δικτύου	40
Πίνακας 6. Μέσος αριθμός Συσχέτισης UE	62
Πίνακας 7. Μέσος αριθμός Συσχέτισης με Handover	63

Λιστα Κομματιών Κώδικα

Κώδικας 1. Κλάση Point	28
Κώδικας 2. Κλάση Σταθμού Βάσης (BS).....	30
Κώδικας 3. Δημιουργία Τυχαίων Σημείων.....	32
Κώδικας 4. Απεικόνιση Σημείων	32
Κώδικας 5. Απεικόνιση BS	34
Κώδικας 6. Κώδικας Δημιουργίας BS (part 1).....	42
Κώδικας 7. Κώδικας Δημιουργίας BS (part 2).....	43

Συντομογραφίες

5G	5th Generation	Mbs	Macro-Base Station
6G	6th Generation	MDP	Markov Decision Process
AI	Artificial Intelligence	ML	Machine Learning
ANN	Artificial Neural Network	mMTC	Massive Machine Type Communication
AoA	Angle Of Arrival	mmWave	Millimeter Wave
API	Application Programming Interface	MPLS	Multiprotocol Label Switching
ASIC	Application Specific Integrated Circuit	MRE	Modified Roth–Erev
BBU	Baseband Unit	MS	Mobile Station
BM	Behavioral Model	NE	Nash Equilibrium
BS	Base Station	NFV	Network Function Virtualization
CNN	Convolutional Neural Network	NN	Neural Network
CP	Control Plane	OF	Open Flow
C-RAN	Cloud Radio Access Network	ONF	Open Network Foundation
CRE	Cell Range Extension	O-RAN	Open-Radio Access Network
CRN	Cognitive Radio Network	OSPF	Open Shortest Path First
CSI	Channel State Information	OVS	Open Vswitch

D2D	Device-To-Device	PBS	Pico-Base Station
DCN	Data Center Network	PCA	Principal Component Analysis
DL	Deep Learning	PISA	Protocol Independent Switch Architecture
DNM	Deep Neural Model	POF	Protocol Oblivious Forwarding
DNN	Deep Neural Network	PU	Primary User
DPDK	Data Plane Development Kit	Pvs	Pof Switch
DPPO	Distributed Proximal Policy Optimization	QoE	Quality Of Experience
DQN	Deep Q-Network	QoS	Quality Of Service
DRL	Deep Reinforcement Learning	RAN	Radio Access Network
DSL	Domain-Specific Language	RB	Resource Block
DT	Decision Tree	RIC	Ran Intelligent Controller
ECR	Energy Consumption Ratio	RL	Reinforcement Learning
ELM	Extreme Learning Machine	RNN	Recurrent Neural Network
eMBB	Enhance Mobile Broadband	RRH	Remote Radio Head
Enb	Evolved Node B	RRU	Remote Radio Unit
EST	Echo State Network	RSRP	Reference Signal Received Power
ETL	Extract-Transform-Load	RSRQ	Reference Signal Received Quality
FBS	Femto-Base Station	RSS	Received Signal Strength

FC	Femtocell	RSSI	Received Signal Strength Indicator
FFNN	Feed-Forward Neural Network	RSU	Roadside Unit
GF	Gradient Follower	SBS	Small Base Station
gNB	Gigabit Node B	SCN	Small Cell Network
GPU	Graphics Processing Unit	SDN	Software-Defined Networking
HAI	Hardware Abstraction Librar	SINR	Signal-To-Interference-Plus-Noise
HetNet	Heterogeneous Network	SL	Supervised Learning
ICIC	Inter-Cell Interference Coordination	SSU	Spectrum Selection Utility
IDE	Integrated Development Environment	SNR	Signal-To-Noise Ratio
NINT	In-Band Network Telemetry	SOM	Self-Organizing Map
IoT	Internet Of Things	SON	Self-Organizing Networks
IP	Internet Protocol	SSU	Spectrum Selection Utility
ISP	Internet Service Provider	SVD	Singular Value Decomposition
IVN	Intelligent Vehicular Networks	SVM	Support Vector Machine
JSON	Javascript Object Notation	SU	Secondary User
KDN	Knowledge-Defined Networking	TL	Transfer Learning
KNN	K-Nearest Neighbors	TTP	Table Type Patterns
KP	Knowledge Plane	UAV	Unmanned Aerial Vehicle

LFU	Least Frequently Used	UDN	Ultra Dense Networks
LoS	Line-Of-Sight	UE	User Equipment
LRU	Least Recently Used	UL	Unsupervised Learning
LTE	Long-Term Evolution	V2I	Vehicle To Infrastructure
LTE-U	Lte-Unlicensed	V2V	Vehicle-To-Vehicle
M2M	Machine-To-Machine	Vanet	Vehicular Ad Hoc Network
MAB	Multi-Armed Bandit	V-RAN	Virtual Radio Access Network
MAC	Medium Access Control	WMMSE	Weighted Minimum Mean Square Error
MBM	Modified Bushand Mostelle	WSN	Wireless Sensor Network

1

ΕΙΣΑΓΩΓΗ

1.1 Προηγούμενες Γενιές Κινητών Δικτύων Επικοινωνίας

Τα τελευταία χρόνια οι κινητές επικοινωνίες έχουν γίνει αναπόσπαστο κομμάτι της καθημερινότητας μας και μια σημαντική πηγή ανταλλαγής πληροφοριών σε όλο τον κόσμο. Όλες οι τεχνολογίες κινητής τηλεφωνίας που έχουν αναπτυχθεί μέχρι σήμερα έχουν βελτιώσει πραγματικά τον τρόπο ζωής μας. Η ασύρματη τεχνολογία βοηθά στη μεταφορά πληροφοριών σε απόσταση χωρίς τη χρήση καλωδίων.

Οι κινητές συσκευές έχουν εμφανιστεί με διαφορετικά πρόσωπα, αν ανατρέξουμε στην ιστορία των κινητών τηλεφώνων, βλέπουμε ότι η πρώτη συσκευή δεν ήταν πραγματικό τηλέφωνο. Στην πραγματικότητα ήταν αμφίδρομοι ασύρματοι, που επέτρεπαν στους χρήστες όπως οι οδηγοί ταξί να επικοινωνούν. Αντί να χρησιμοποιούν Σταθμούς Βάσης με ξεχωριστές κυψέλες και το σήμα να περνάει μεταξύ των κυψελών, το πρώτο δίκτυο κινητής τηλεφωνίας βασίστηκε σε έναν ισχυρό Σταθμό Βάσης που επέτρεπε επικοινωνία σε πολύ ευρύτερη περιοχή.[9]

Τα δίκτυα επικοινωνίας έχουν εξελιχθεί μέσα από τρία μεγάλα άλματα γενεών ακολουθώντας τις τάσεις της τεχνολογίας και τις διαρκώς εξελισσόμενες απαιτήσεις των χρηστών. Το πρώτο εξελικτικό άλμα ήταν από την πρώτη γενιά, γνωστή ως 1G, στη δεύτερη γενιά, δηλαδή στο 2G, όταν ψηφιοποιήθηκε το δίκτυο κινητής φωνής. Το επόμενο εξελικτικό άλμα από το 2G στο 3G έγινε προκειμένου να ικανοποιηθεί η συνεχώς αυξανόμενη ζήτηση των χρηστών για δεδομένα και ποιότητα υπηρεσιών (Quality of Service - QoS).

Ο πολλαπλασιασμός εξελιγμένων πλατφορμών χρηστών, όπως τα έξυπνα τηλέφωνα και τα tablet, και η ανάπτυξη νέων εφαρμογών για κινητά με ένταση εύρους ζώνης τροφοδότησε περαιτέρω την απαίτηση των χρηστών για εύρος ζώνης και ποιότητα. Αυτό οδήγησε στο επόμενο εξελικτικό άλμα προς το 4G, το οποίο έκανε τα δίκτυα κινητής τηλεφωνίας να παρέχουν μια πραγματική ασύρματη ευρυζωνική υπηρεσία στους πελάτες τους. Με τις βελτιωμένες επιλογές που προσφέρει το 4G, εισήχθησαν νέες περιπτώσεις χρήσης, όπως στους τομείς της υγείας, της αυτοκινητοβιομηχανίας, της ψυχαγωγίας, της βιομηχανίας, της κοινωνικής και του περιβάλλοντος με ποικίλες απαιτήσεις υπηρεσιών. Οι υπηρεσίες καινοτομούν γρήγορα με υπερβολική εξάρτηση από την υποδομή του δικτύου κινητής τηλεφωνίας για τις ανάγκες συνδεσιμότητας τους. Με μια τέτοια εξέλιξη και το Διαδίκτυο να μετατρέπεται σε ένα Διαδίκτυο των Πραγμάτων (Internet of Things – IoT), η έννοια του πελάτη

έχει αλλάξει από ανθρώπινους πελάτες μέχρι τώρα να περιλαμβάνει επίσης αυτοκίνητα, αισθητήρες, ηλεκτρονικά είδη ευρείας κατανάλωσης, μετρητές ενέργειας κ.λπ.

Mobile Generation	GSM(2G)	CDMA(3G)	LTE(4G)	5G
Refers to	Second generation	Third generation	Fourth generation	Fifth generation
Deployment	1990	2001	2010	2020
Features	It uses digital signals. It provides services such as text messages etc	It provides faster communication. It has large capacities and broadband capabilities. It send/receive large email messages	provides additional services. It provides mobile ultrabroadband Internet access.	Interactive multimedia, voice internet and others services as supported by 5G. It is more effective and attractive as compare to other generations
Band width	200 KHz	5 MHz	20 MHz	Up to 100 MHz
Data rate	14.4 Kbps	3.1 Mbps	100 Mbps	20 Gbps- 100 Gbps
Latency	700 ms	< 200 ms	< 30 ms	About 1 ms
Modulation Technique	GMSK	QPSK, 16QAM	QPSK, 16 QAM and 64 QAM	256 QAM
Applications	Voice and Slow data rate communication	Advanced applications(various services like data services access to television/video)	high rate data Applications, wearable devices	device-to-device, machine-to-machine, internet of Things ,2way gaming, virtual reality glasses, cloud-based computing and other technologies
Multiplexing	TDMA	W-CDMA	OFDMA and MIMO	OFDM, NOMA, FBMC Massive MIMO
Standards	GSM,EDGE,GPRS	UMTS, CDMA2000, HSPDA,EVDO	LTE advanced, IEEE 802.16 (WiMAX)	CDMA,BDMA
Switching	circuit/packet	Packet	all packet	all packet
Hand over	Horizontal	Horizontal	Horizontal/vertical	Horizontal/vertical
Core type Network	PSTN	Packet- Network	Internet	Internet
Technology	Digital cellular	code division multiple access, universal mobile Telecommunications system	Long-Term evolution advanced, Wi-Fi	Multi- radio Access technology, Wi-Fi, Wi-Gig

Εικόνα 1 Συγκριτικός πίνακας προηγούμενων γενιών

Με μια τόσο ποικιλόμορφη βάση πελατών, το δίκτυο κινητής τηλεφωνίας όχι μόνο πρέπει να διαχειρίζεται τον αυξανόμενο όγκο δεδομένων, αλλά ταυτόχρονα να διασφαλίζει ότι τα αιτήματα εξυπηρέτησης πελατών ικανοποιούνται επαρκώς από το δίκτυο, ικανοποιώντας την αντίστοιχη ποιότητα υπηρεσίας ή τις απαιτήσεις σε ποιότητα εμπειρίας. Προκειμένου να ανταποκριθούν στις απαιτήσεις δεδομένων και υπηρεσιών, οι φορείς εκμετάλλευσης δικτύων επεκτείνουν και αναβαθμίζουν συνεχώς την υποδομή του δικτύου τους, με αποτέλεσμα αυξημένες κεφαλαιακές και λειτουργικές δαπάνες. Ωστόσο, παρά την παρουσία του έντονου ανταγωνισμού σε συνδυασμό με την πτώση των τιμών, τα μέσα έσοδα ανά χρήστη δεν αυξάνονται αναλογικά με αποτέλεσμα η επένδυση από τους παρόχους δικτύων να μην επιφέρει τους αναμενόμενους καρπούς. Έτσι, προκειμένου να μειωθεί το κόστος και να αυξηθούν τα

έσοδα, τα δίκτυα κινητής τηλεφωνίας πρέπει να κάνουν το επόμενο εξελικτικό τους άλμα προς το 5G, το οποίο πλέον δεν απευθύνεται μόνο στο πλεονέκτημα της κινητής τηλεφωνίας, αλλά και στο κεντρικό δίκτυο.[11]

1.2 Δίκτυα 5ης Γενιάς

Τα δίκτυα 5G, που αναφέρονται επίσης ως συστήματα επικοινωνιών μετά το 2020, αντιπροσωπεύουν την επόμενη σημαντική φάση της τηλεπικοινωνιακής βιομηχανίας. Ανάμεσα στα πολυάριθμα χαρακτηριστικά των δικτύων 5G, τα τρία κυριότερα που διακρίνουμε είναι πως το δίκτυο 5G θα έχει την ικανότητά του να υποστηρίζει Βελτιωμένη MobileBroadband, μαζική επικοινωνία τύπου μηχανής και να παρέχει εξαιρετικά αξιόπιστες υπηρεσίες επικοινωνίας χαμηλής καθυστέρησης. Αυτό συνεπάγεται ότι τα δίκτυα 5G σε ιδανικές συνθήκες παρέχουν αυξημένους ρυθμούς bit σε Gbps ανά χρήστη, υψηλότερη απόδοση φάσματος, καλύτερη κάλυψη και υποστήριξη για μαζικά αυξημένο αριθμό διαφορετικών συνδεδεμένων συσκευών. Επιπλέον, τα συστήματα 5G απαιτείται να είναι οικονομικά αποδοτικά, αξιόπιστα, ευέλικτα στην ανάπτυξη, ελαστικά, ευέλικτα και κυρίως προγραμματιζόμενα. Αυτές είναι φιλόδοξες και εξαιρετικά απαιτητικές απαιτήσεις που έχουν επιπτώσεις τόσο στο δίκτυο κινητής ραδιοπρόσβασης όσο και στο δίκτυο κινητής τηλεφωνίας, και επομένως απαιτούν σημαντικό επανασχεδιασμό και ανασχεδιασμό τόσο της αρχιτεκτονικής όσο και της τεχνολογίας. Νέες καινοτόμοι μέθοδοι και συστήματα διερευνώνται και αξιολογούνται προκειμένου να επιτευχθούν οι απαιτητικοί στόχοι απόδοσης των δικτύων 5G.

1.2.1 Αρχιτεκτονική Δικτύων 5ης Γενιάς

Φυσικό Επίπεδο και Συχνοτικές Ζώνες

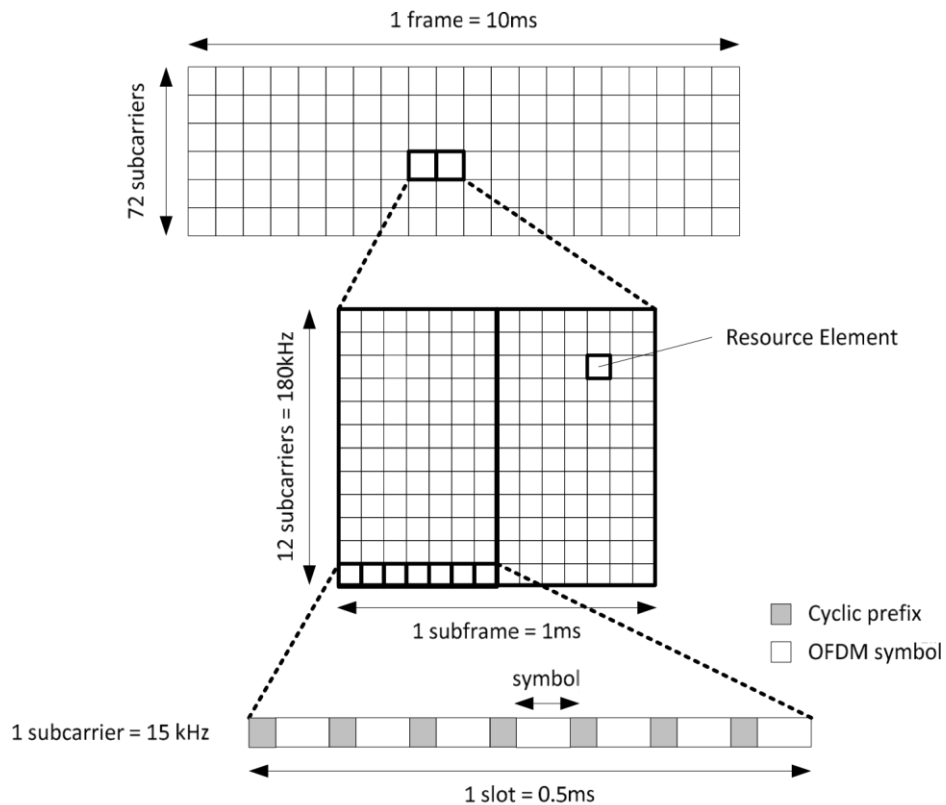
Υπάρχουν δύο κύρια στοιχεία στο δίκτυο 5G NR. Το πρώτο από αυτά είναι το User Equipment - UE (δηλαδή συνδρομητής κινητής τηλεφωνίας) και gNB (δηλαδή σταθμός βάσης). Τα gNB συνδέονται με 5G Core στο backend. Η σύνδεση από το gNB στο UE είναι γνωστή ως κατερχόμενη ζεύξη (Downlink) που χρησιμοποιεί κανάλια PBCH, PDSCH και PDCCH για τη μεταφορά διαφορετικών δεδομένων/πληροφοριών ελέγχου. Η σύνδεση από το UE στο gNB είναι γνωστή ως ανερχόμενη ζεύξη (Uplink) που χρησιμοποιεί κανάλια PRACH, PUSCH και PUCCH.

Στο 5G NR χρησιμοποιούνται 2 εύρη συχνότητων. Το πρώτο είναι το FR1 για συχνότητες μικρότερες των 6GHz με Time Division Duplex (TDD) Frequency Division Duplex (FDD) και FR2 για συχνότητες μεταξύ 24 – 52.6 GHz μόνο με TDD.

Η απόσταση μεταξύ δύο subcarriers σε 5G είναι μεταβλητή και είναι πολλαπλάσια του 15 KHz μέχρι τα 240 KHz. Ανάλογα με τα εύρη, το μέγιστο εύρος ζώνης (bandwidth) και η απόσταση φορέων (Subcarrier Spacing – SCS) ποικίλλει. Στο εύρος συχνοτήτων υπό των 6 GHz, το μέγιστο bandwidth είναι 100 Mhz και στο εύρος mmWave το μέγιστο bandwidth είναι 400 Mhz. Κάποιες αποστάσεις υποφορέων (15, 30 KHz) μπορούν να χρησιμοποιηθούν μόνο σε Sub 6 GHz και κάποιες αποστάσεις υποφορέων (120 KHz) μπορούν να χρησιμοποιηθούν μόνο σε

mmWave, και κάποιες αποστάσεις υποφορέων (60 KHz) μπορούν να χρησιμοποιηθούν και στις δύο υπό 6 GHz και mmWave.

Έτσι έχουμε ως αποτέλεσμα η διάρκεια ενός OFDM συμβόλου να είναι αντίστροφη της απόστασης μεταξύ των subcarriers και υποδιαιρείται με 2, 4, 8, 16 στην μεγαλύτερη απόσταση δύο subcarrier. Το μικρότερο υποστηριζόμενο bandwidth είναι 50 MHz στα 15 KHz SCS και διπλασιάζεται κάθε φορά μέχρι να φτάσει τα 100, 200, 400 MHz στα 120 KHz spacing. Για την υψηλότερη απόσταση μεταξύ δύο subcarriers στα 240 kHz, ο αριθμός των υποστηριζόμενων subcarriers μειώνεται στο μισό, πράγμα που σημαίνει ότι το εύρος ζώνης εξακολουθεί να είναι 400 MHz. Καθώς η απόσταση των subcarriers αυξάνεται, το υποστηριζόμενο εύρος ζώνης αυξάνεται και η διάρκεια του συμβόλου OFDM και οι καθυστερήσεις στο φυσικό επίπεδο μειώνονται. Το 5G NR διατηρεί την έννοια του πλαισίου (frame) 10 msec χωρισμένο σε 10 υποπλαίσια διάρκειας 1 msec το καθένα, ωστόσο, ο ορισμός της υποδοχής είναι διαφορετικός από το LTE.[40]



Εικόνα 2. Resource Grid

1.2.2 Τεχνολογίες Δικτύων 5ης Γενιάς

Ετερογενή Δίκτυα (HetNets)

Τα HetNets δημιουργήθηκαν με στόχο να επεκτείνουν τις προϋπάρχουσες υποδομές των πιο διαδεδομένων μορφών κυψελίδων που είναι τα Macrocells. Αυτό επιτυγχάνεται εγκαθιστώντας μικρά κελιά εντός της εμβέλειας των Macrocells σε συγκεκριμένες τοποθεσίες, όπως κοντά στα όρια των Macrocells. Με αυτό τον τρόπο μπορούμε να παρέχουμε καλύτερη κάλυψη και απόδοση στα UE κοντά στα όρια του Σταθμού Βάσης (BS) όπου σε αυτό το σημείο είναι πιο ισχυρά τα επίπεδα παρεμβολών. Στα HetNets τα BS υπάρχουν όλο και πιο κοντά στα UE δημιουργώντας πιο πυκνά BSs με αποτέλεσμα να πετυχαίνουμε υψηλότερη φασματική απόδοση και καλύτερη ενεργειακή κατανομή για τα κυψελωτά συστήματα.

Απόξευξη Κάτω και Άνω Ζεύξης (Downlink and Uplink Decoupling – DUDe)

Το DUDe επιτρέπει στον Εξοπλισμό Χρήστη (UE) να συνδεθεί σε διαφορετικό BS για κάθε διεύθυνση DL και UL. Σε ένα δίκτυο που χρησιμοποιεί την τεχνολογία DUDe, δύο λίστες (μία για διευθύνσεις DL και μία για διευθύνσεις UL) που αποτελούνται από BSs που μπορούν να χρησιμοποιηθούν για συσχέτιση δημιουργούνται για κάθε UE. Κάθε λίστα ταξινομείται σύμφωνα με δείκτες όπως ο λόγος σήματος προς παρεμβολή συν Θόρυβος (SINR). Μετά την προετοιμασία της λίστας, κάθε UE προσπαθεί να συνδεθεί στο προτιμώμενο BS. Εάν το BS μπορεί να παρέχει το ζητούμενο Μπλοκ πόρων (RB), δημιουργείται μια σύνδεση μεταξύ των δύο. Εάν δεν υπάρχουν αρκετά RB, το UE προσπαθεί να συνδεθεί στο επόμενο προτιμώμενο BS. Η παραπάνω διαδικασία εκτελείται και προς τις δύο κατευθύνσεις έως ότου κάθε UE συνδεθεί σε ένα BS. Η κύρια ιδέα του DUDe είναι ότι αντιμετωπίζει τις διευθύνσεις DL και UL ως δύο διαφορετικές εργασίες, που οδηγούν σε ταχύτερες συσχετίσεις και καλύτερη εμπειρία χρήστη.[8]

Διαμόρφωση Δέσμης (Beamforming)

Υπάρχουν 3 τύποι διαμόρφωσης δέσμης (BF), το στατικό BF εκτελείται χρησιμοποιώντας πολλές κατευθυντικές κεραιές ομαδοποιημένες με στόχο μακριά από ένα κεντρικό σημείο για να παρέχουν ένα σταθερό σχέδιο ακτινοβολίας. Το Dynamic BF μπορεί να προσαρμόσει δυναμικά το μοτίβο ακτινοβολίας για να παρέχει το καλύτερο σήμα για κάθε συσκευή που συνδέεται. Η προσαρμοστική διάταξη κεραιάς χρησιμοποιείται σε αυτόν τον τύπο για ελιγμούς της δέσμης προς την κατεύθυνση ενός στοχευμένου Δέκτη και αυτή η τεχνολογία ονομάζεται «τεχνολογία έξυπνης κεραιάς» ή «διεύθυνση δέσμης». Μπορεί να εστιάσει μια δέσμη προς την κατεύθυνση ενός μεμονωμένου χρήστη. Τέλος, η διαμόρφωση δέσμης μετάδοσης (TxBF) εκτελείται με τη μετάδοση πολλαπλών σημάτων μετατόπισης φάσης και όλα αυτά τα σήματα ελπίζουμε ότι θα φτάσουν στη φάση του δέκτη.[10]

5G Slicing

Οι απαιτήσεις των δικτύων 5G που συζητήθηκαν παραπάνω οδηγούν σε μια πολύ σημαντική ιδέα της τεμαχίσεως (slicing) που έχει γίνει κεντρικό θέμα στα δίκτυα 5G. Ο τεμαχισμός δικτύου επιτρέπει στους χειριστές δικτύου να ανοίγουν την φυσική τους πλατφόρμα υποδομής δικτύου στην ταυτόχρονη ανάπτυξη πολλαπλών λογικών αυτοτελών δικτύων, ενορχηστρωμένων με διαφορετικούς τρόπους σύμφωνα με τις ειδικές απαιτήσεις υπηρεσιών

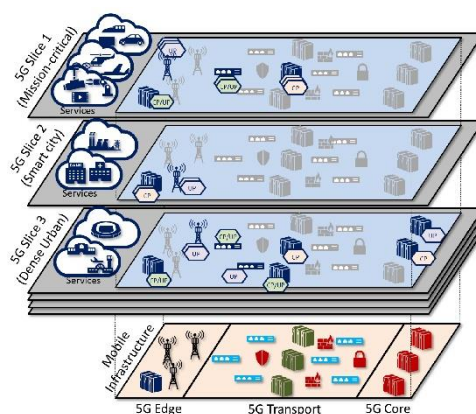
τους. Αυτά τα τμήματα δικτύου στη συνέχεια αποκτούνται (προσωρινά) σε ενοικιαστές. Καθώς αυτοί οι ενοικιαστές έχουν τον έλεγχο πολλαπλών επιπέδων, όπως το φυσικό επίπεδο, το επίπεδο εικονικοποίησης και το επίπεδο υπηρεσιών, μιας υποδομής 5G, ονομάζονται επίσης κάθετες, δηλαδή ενσωματώνουν την υποδομή 5G κάθετα. Η διαθεσιμότητα αυτής της κάθετης αγοράς πολλαπλασιάζει τις ευκαιρίες δημιουργίας εσόδων της υποδομής δικτύου καθώς νέες επιχειρήσεις, όπως η αυτοκινητοβιομηχανία και η ηλεκτρονική υγεία, ενδέχεται να μπου στην αγορά, και επειδή μπορεί να επιτευχθεί υψηλότερη αξιοποίηση της χωρητικότητας υποδομής με την έταξη του τεμαχίου δικτύου και αξιοποίηση κερδών πολυπλεξίας. Με τον τεμαχισμό δικτύου, διαφορετικές υπηρεσίες, Enhanced Mobile Broadband (eMBB), Ultra-reliable and Low Latency Communications (URLLC), Massive Machine-Type Communications (mMTC). Κάθε μία από αυτές τις περιπτώσεις αποτελείται από ένα σύνολο λειτουργιών εικονικού δικτύου (Virtual Network Functions – VNFs) που εκτελούνται στην ίδια υποδομή με προσαρμοσμένη ενορχηστροποίηση. Με αυτόν τον τρόπο, μπορούν να παρέχονται διάφορες ετερογενείς απαιτήσεις από την ίδια υποδομή, καθώς διαφορετικά στιγμιότυπα τμημάτων δικτύου μπορούν να ενορχηστρωθούν και να διαμορφωθούν ξεχωριστά σύμφωνα με τις συγκεκριμένες απαιτήσεις. Επιπλέον, αυτό εκτελείται με τρόπο αποδοτικό ως προς το κόστος, καθώς οι διαφορετικοί ενοικιαστές τμημάτων δικτύου μοιράζονται την ίδια φυσική υποδομή.

Το Network Slice ορίζεται από το NGMN (Next Generation Mobile Networks) ως ένα σύνολο συναρτήσεων δικτύου και πόρων για την εκτέλεση των λειτουργιών δικτύου, σχηματίζοντας ένα πλήρες λογικό δίκτυο για την κάλυψη ορισμένων χαρακτηριστικών δικτύου που απαιτούνται από τις απαιτούμενες υπηρεσίες. Σύμφωνα με το NGMN, η έννοια του τεμαχισμού δικτύου περιλαμβάνει τρία επίπεδα:

- Επίπεδο Απαιτούμενων Υπηρεσιών
- Επίπεδο Τεμαχίου Δικτύου
- Επίπεδο Πόρων

Το επίπεδο Απαιτούμενων Υπηρεσιών αντιπροσωπεύει τις υπηρεσίες του τελικού χρήστη ή/και των επιχειρήσεων, που παρέχονται από τον χειριστή ή τους παρόχους υπηρεσιών τρίτου μέρους, οι οποίες υποστηρίζονται από το επίπεδο Τεμαχίου Δικτύου. Το επίπεδο Τεμαχίου Δικτύου υποστηρίζεται με τη σειρά του από το επίπεδο Πόρων, το οποίο μπορεί να αποτελείται από φυσικούς πόρους όπως υπολογιστές, δίκτυο, μνήμη, αποθήκευση κ.λπ. ή μπορεί να είναι πιο ολοκληρωμένο ως υποδομή δικτύου ή μπορεί να είναι πιο περίπλοκο ως λειτουργίες δικτύου.

Στην εικόνα 2 απεικονίζεται αυτή η ιδέα όπου οι πόροι στα επίπεδα πηγών έχουν διαστάσεις για τη δημιουργία πολλών περιπτώσεων υποδικτύου και σχηματίζονται στιγμιότυπα από network slices που μπορούν να χρησιμοποιούν κανένα, ένα ή πολλά στιγμιότυπα υποδικτύου.



Εικόνα 3. Network Slicing

Λαμβάνοντας υπόψη τις παραπάνω προκλήσεις, έχουν αναπτυχθεί δύο βασικές τεχνολογίες για την κάλυψη των απαιτήσεων επεκτασιμότητας, ευελιξίας, και προγραμματισμού των δικτύων κινητής τηλεφωνίας 5G, η Εικονικοποίηση λειτουργιών δικτύου (Network Function Virtualization - NFV) και Δικτύωση Βασισμένη στο Λογισμικό (Software Defined Networking - SDN).[14]

Εικονικοποίηση λειτουργιών δικτύου (Network Function Virtualization - NFV)

Τα NFV συστήματα στοχεύουν στην μείωση του κόστους όσο αφορά την ένταξη νέων υπηρεσιών, συντήρηση των υπάρχων υπηρεσιών και διαχείριση των διαθέσιμων πόρων. Στόχος της τεχνολογίας NFV είναι η προσομοίωση λειτουργιών που κανονικά θα έτρεχαν σε εξειδικευμένο υλικό ενώ τώρα δίνεται η δυνατότητα προσομοίωσης τους σε virtual machines που τρέχουν σε cloud υποδομές. Η πρώτη γενιά NFV συστημάτων μετέφερε παλιές εφαρμογές σε προσομοιωτές όπου κάθε ένας από αυτούς τους προσομοιωτές αποτελούσε ένα VFN. Στην συνέχεια πολλά VFN δικτυώνονται για να δημιουργήσουν ενός Service Function Chain που καθορίζει πως τα πακέτα προωθούνται από ένα VFN σε άλλο, και μαζί αποτελούν μια υπηρεσία δικτύου. Ωστόσο αποδείχθηκε ότι αυτή η πρακτική δεν μεταβιβάζεται αποδοτικά σε cloud. Πύλες λογισμικού σχεδιασμένες για να τρέχουν σε εξειδικευμένο υλικό όπως προαναφέραμε αδυνατούν να αποδώσουν σε περιβάλλοντα cloud. Έτσι οι σημερινές προσεγγίσεις αποσκοπούν στη σχεδίαση λογισμικού για εφαρμογές δικτύου που να αποδίδει σε cloud.

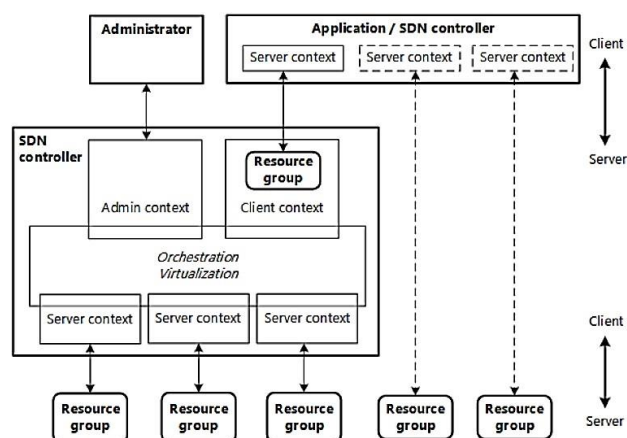
Αυτό όμως παρουσιάζει προβλήματα στην διαχείριση NFV καθώς αυτός ο τρόπος σχεδίασης λογισμικού συνεπάγεται την ανάγκη για περισσότερες NFV οντότητες που χρειάζεται να ενορχηστρωθούν. Για αυτό τον λόγο έχουν αναπτυχθεί τα MANO συστήματα με σκοπό την διαχείριση της εικονοποιημένης και δικτυακής δομής. Ένα 5G δίκτυο στον πυρήνα του αποτελείται από 3 επίπεδα όπου κάθε ένα από αυτά τα επίπεδα χρειάζεται να χειρίζεται σε συνδυασμό με τα υπόλοιπα και αυτό είναι που επιχειρούν να κάνουν τα NFV MANO συστήματα. Το πιο γνωστό από αυτά είναι το ETSI MANO και αποτελείται από 3 blocks. Το πρώτο είναι το NFVO (Network Function Virtualization Orchestrator) που χειρίζεται τις δικτυακές υπηρεσίες, το VNF Manager που είναι υπεύθυνο για τον κύκλο ζωής μιας λειτουργίας εικονικού δικτύου προσφέροντας σημεία αναφοράς για σύνδεση με legacy

συστήματα και το VIM (Virtualized Infrastructure Manager) που συνδέεται με NFV υποδομές και βοηθάει στην διαχείριση των λειτουργιών στο επίπεδο των εικονικών μηχανών αλλά και στην συνδεσιμότητα των διαφόρων VNFs. Κάποιες επιπλέον λειτουργίες που προσφέρει η ETSI πέρα του MANO framework είναι το NSD (Network Service Descriptor) που παρέχει πληροφορίες για τα VNFs.

Δικτύωση Βασισμένη στο Λογισμικό (Software Defined Networking - SDN)

Το SDN δημιουργήθηκε για να ξεπεραστούν οι περιορισμοί των παραδοσιακών δικτύων και των παραδοσιακών λειτουργιών δικτύου. Το κίνητρο ήταν η διάσπαση των αποστεωμένων δικτύων αντικαθιστώντας άκαμπτο ιδιόκτητο εξοπλισμό και υπηρεσίες που βασίζονται σε υλικό, με βαθιά προγραμματιζόμενες υπηρεσίες και μεθόδους που βασίζονται σε λογισμικό και εκτείνονται σε πολλαπλές πλατφόρμες προμηθευτών. Διαχωρίζοντας έτσι τους κύκλους του εύελικτου λογισμικού από τους αργούς κύκλους του ενσωματωμένου λογισμικού και υλικού. Η ριζική αλλαγή προς το να γίνουν τα δίκτυα προγραμματιζόμενα, και οι εφαρμογές και οι δικτυακές υπηρεσίες να ελέγχουν άμεσα την αφηρημένη υποδομή, πυροδότησε μια νέα αναπτυξιακή κατεύθυνση στα ερευνητικά και εκπαιδευτικά δίκτυα και στα εμπορικά δίκτυα, επηρεάζοντας ιδιαίτερα εδραιωμένους προμηθευτές δικτυακού εξοπλισμού.

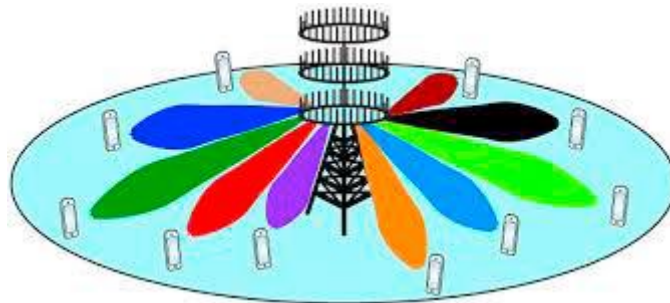
Το SDN έχει κερδίσει μεγάλη έλξη τα τελευταία χρόνια. Η δυνατότητα διαχείρισης υπηρεσιών δικτύου μέσω αφαιρέσεων λειτουργιών χαμηλότερου επιπέδου ανοίγει ένα ευρύ φάσμα επιλογών νέας αρχιτεκτονικής, διαχείρισης και λειτουργίας, συμπεριλαμβανομένων νέων μορφών αλληλεπίδρασης μεταξύ εφαρμογών τελικού χρήστη και δικτύων. Η ανάπτυξη εύελικτου λογισμικού σχεδιασμένο για μεταγωγείς δικτύου αναμένεται να βελτιώσει τη συμπεριφορά κόστους-απόδοσης του δικτύου. Μια άλλη μεγάλη αξία του SDN θα είναι η δυνατότητα ταχείας παράδοσης των υπηρεσιών χρήστη, ενώ παράλληλα θα χρησιμοποιούνται οι πόροι του δικτύου πιο αποτελεσματικά.



Εικόνα 4. Η Αρχιτεκτονική ενός SDN Συστήματος

Massive MIMO

Το Massive MIMO προτάθηκε για πρώτη φορά από τον Thomas Marzetta από τα Bell Labs το 2010. Είναι μια επεκτάσιμη μορφή του MIMO πολλαπλών χρηστών όπου υπάρχουν αρκετές βασικές διαφορές μεταξύ του Massive MIMO και του MIMO πολλαπλών χρηστών. Η προτεινόμενη ιδέα του τεράστιου MIMO είναι να εξοπλίσει το Σταθμό Βάσης με εκατοντάδες κεραιές, οι οποίες είναι πολύ μεγαλύτερες από τον αριθμό των χρηστών. Η διάταξη της κεραιάς λαμβάνει σήματα δεδομένων από τα τερματικά χρήστη και στέλνει επιλεκτικά ροές δεδομένων που μοιράζονται το ίδιο πόρους χρόνου και συχνότητας. Στην κατερχόμενη ζεύξη, κάθε χρήστης θα πρέπει να λαμβάνει μόνο τη ροή δεδομένων που προορίζεται για αυτόν. Ωστόσο, στην ανερχόμενη ζεύξη, ο Σταθμός Βάσης (BS) λαμβάνει και ανακτά τα σήματα δεδομένων που στάλθηκαν από τα τερματικά χρήστη. Ως αποτέλεσμα των συνθηκών οπτικής επαφής, το BS υποστηρίζει μεμονωμένη δέσμη για κάθε ακροδέκτη χρήστη όπως φαίνεται στο. Αυξάνοντας τον αριθμό των κεραιών στο σταθμό βάσης, οι δέσμες που κατευθύνονται στο τερματικό χρήστη έχουν πολύ μικρότερα πλάτη δέσμης και πολύ εστιασμένη ισχύ.



Εικόνα 5. Massive MIMO Network

2

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2.1 Τεχνικές Μηχανικής Μάθησης

Η Μηχανική Μάθηση αποτελεί ένα κομμάτι της Τεχνητής Νοημοσύνης. Στόχος της είναι η μελέτη τεχνικών μάθησης που εκπαιδεύουν μοντέλα ώστε να μπορούν, ανάλογα με την περίπτωση, να κάνουν προβλέψεις και κατηγοριοποιήσεις έχοντας εκπαιδευτεί από δεδομένα του παρελθόντος. Λαμβάνοντας υπόψη τον τρόπο μάθησης, και το είδος των δεδομένων που διαχειριζόμαστε, μπορούμε να κατηγοριοποιήσουμε τις τεχνικές της ΜΜ σε 3 κατηγορίες:

- Μάθηση με Επίβλεψη: Γίνεται χρήση κατηγοριοποιημένων δεδομένων που έχουν γνωστές επιθυμητές εισόδους και εξόδους.
- Μάθηση χωρίς Επίβλεψη: Η μάθηση χωρίς επίβλεψη δεν απαιτεί κατηγοριοποιημένα δεδομένα για την φάση της εκπαίδευσης και έτσι απαιτεί μόνο δεδομένα εισόδου χωρίς επιθυμητή έξοδο που να εμπίπτει σε κάποια κατηγορία, ο ίδιος ο αλγόριθμος θα παράξει τις κατηγορίες στις οποίες εμπίπτουν τα δεδομένα σύμφωνα με τις ομοιότητες που φέρουν.
- Ενισχυτική Μάθηση: Επιτρέπει την μάθηση από την ανατροφοδότηση δεδομένων που λαμβάνονται μέσω των αλληλεπιδράσεων με ένα εξωτερικό περιβάλλον. Δεν δίνονται οδηγίες στον πράκτορα ως το προς ποια δράση πρέπει να πάρει κάθε φορά, αλλά πρέπει να ανακαλύψει ποιο σύνολο δράσεων θα επιφέρει το βέλτιστο αποτέλεσμα δοκιμάζοντας κάθε δράση σε κάθε γύρο.

Η Μάθηση με Επίβλεψη και χωρίς Επίβλεψη εστιάζουν κυρίως στην ανάλυση δεδομένων ενώ η Ενισχυτική μάθηση στοχεύει στην λήψη αποφάσεων.

Η ΜΜ βρίσκει εφαρμογές σε διάφορους τομείς. Χρησιμοποιώντας αλγόριθμους κατηγοριοποίησης μπορούμε να ανιχνεύσουμε εξαπάτηση, να αναγνωρίσουμε τι απεικονίζεται

σε μια εικόνα. Χρησιμοποιώντας αλγόριθμους Παλινδρόμησης μπορούμε να κάνουμε προβλέψεις και προγνώσεις σειρών.[4]

Methods	Classification	Features	Representative Algorithm	
Supervised Learning	Decision Tree	Decision Models	ID3, C4.5, CART, CHAID, M5	
	Linear Model	Regression	Ordinary Least Squares Regression, OLER	
			Linear Regression	Ridge Regression LASSO Regression
			Logistic Regression	
			Locally Estimated Scatterplot Smoothing, LOESS	
			Multivariate Adaptive Regression Splines, MARS	
			Classification	Perceptron Model, PM
	Support Vector Machine, SVM			
	Logistic Regression, LR			
	Neural Networks	Auto Encoder, AE		
		Restricted Boltzmann Machine, RBM		
		Artificial Neural Networks, ANN		
Convolutional Neural Networks, CNN				
Recurrent Neural Networks, RNN				
Long Short-Term Memory, LSTM				
Generative Adversarial Networks, GAN				
Deep Learning				
k-Nearest Neighbor, kNN	Distance Metric			
Bayesian Algorithms	Naive Bayes	Naive Bayes Classifier Bayesian Network		
	Normal Bayes	Normal Bayes Classifier		
Unsupervised Learning	Dimensionality Reduction	Linear Dimensionality Reduction	Principal Component Analysis, PCA	
		Nonlinear Dimensionality Reduction	Manifold Learning, ML	
			Laplacian Eigenmaps, LE	
	Local Linear Embedding, LLE			
	Clustering	Hierarchical distance	AGglomerative NESTing , AGNES	
		Prototype-based Clustering	k-means Clustering	
			Learning Vector Quantization, LVQ	
		Density-based Clustering	Density-Based Spatial Clustering of Applications with Noise, DBSCAN	
			Ordering Points to Tdentify the Clustering Structure , OPTICS	
	Mean shift Algorithm			
Probabilitybased Clustering	Expectation-Maximization Algorithm, EM			
Graph-based Clustering	Spectral Clustering			
Reinforcement Learning	Policy-based Reinforcement Learning	Dynamic Programming	Policy Iteration	
			Value Iteration	
	Value-based Reinforcement Learning	Random Algorithm	Monte Carlo Method	
			Temporal Difference learning, TD Learning	Sarsa Q-learning
	Deep Reinforcement Learning	Reinforcement Learning + Deep Learning	Deep Q-learning, DQN Policy Gradient Algorithm	
Other Methods	Semi-supervised Learning, Ensemble Learning...			

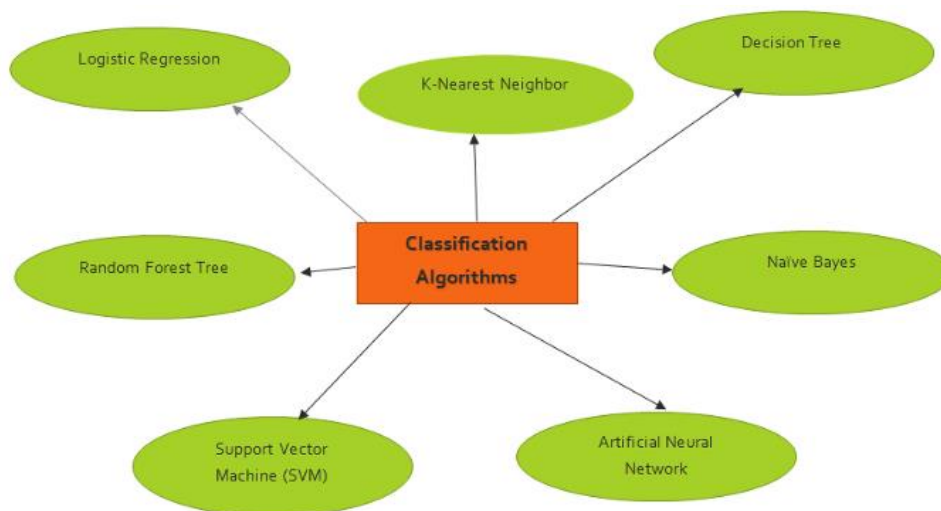
Εικόνα 6. Κατηγορίες Μηχανικής Μάθησης

2.2 Μάθηση με Επίβλεψη

Μπορούμε να διακρίνουμε δύο μεγάλους υποκλάδους στην περίπτωση της Επιβλεπόμενης Μάθησης. Την διαδικασία απεικόνισης γνωστών εισόδων σε γνωστές επιθυμητές εξόδους (ετικέτες), αλλιώς γνωστή και ως Κατηγοριοποίηση, και την διαδικασία εκπαίδευσης ενός μοντέλου σε δεδομένα του παρελθόντος με στόχο την πρόβλεψη έκβασης ενός φαινομένου σε μελλοντικές στιγμές δοθέντος δεδομένων που εμπίπτουν στο ίδιο σύνολο με το οποίο εκπαιδεύτηκε το μοντέλο.

2.2.1 Κατηγοριοποίηση

Αναφερόμαστε στην διαδικασία δημιουργίας και εκπαίδευσης ενός μοντέλου για την πρόβλεψη της ετικέτας κλάσης στην οποία ανήκει το κάθε στιγμιότυπο εισόδου. Μερικές από τις χρήσεις της κατηγοριοποίησης είναι η αναγνώριση χειρόγραφων χαρακτήρων, η αναγνώριση αν κάποιο email είναι spam ή χρήσιμο, σε τομείς όπως η αναγνώριση ομιλίας, αναγνώριση προσώπου, φιλτράρισμα ανεπιθύμητης ηλεκτρονικής αλληλογραφίας, κατηγοριοποίηση εγγράφων, αναγνώριση μοτίβων, αστοχία μηχανών, αναγνώριση παθήσεων στην βιοατρική και πολλά άλλα. Στην συνέχεια θα αναλύσουμε τους αλγόριθμους που χρησιμοποιήθηκαν στην παρούσα διπλωματική κατά το πειραματικό κομμάτι.



Εικόνα 7. Αλγόριθμοι Κατηγοριοποίησης

2.2.2 Παλινδρόμηση

Linear Regression

Είναι μια στατιστική μέθοδος που χρησιμοποιείται για την προγνωστική ανάλυση. Η γραμμική παλινδρόμηση κάνει προβλέψεις για συνεχείς / πραγματικές ή αριθμητικές μεταβλητές όπως πωλήσεις, μισθός, ηλικία, τιμή προϊόντος κ.λ.π.

Ο αλγόριθμος γραμμικής παλινδρόμησης δείχνει μια γραμμική σχέση μεταξύ **εξαρτώμενου** (y) και μίας ή περισσότερων **ανεξάρτητων** (x) μεταβλητών, οπότε καλείται γραμμική παλινδρόμηση. Η γραμμική παλινδρόμηση δείχνει τη γραμμική σχέση, που σημαίνει ότι βρίσκει πώς η τιμή της εξαρτημένης μεταβλητής αλλάζει ανάλογα με την τιμή της ανεξάρτητης μεταβλητής.

Το μοντέλο γραμμικής παλινδρόμησης παρέχει μια κεκλιμένη ευθεία γραμμή που αντιπροσωπεύει τη σχέση μεταξύ των μεταβλητών.

Μαθηματικά εκφράζεται ως εξής: $y = a_0 + a_1 * x + \varepsilon$ όπου:

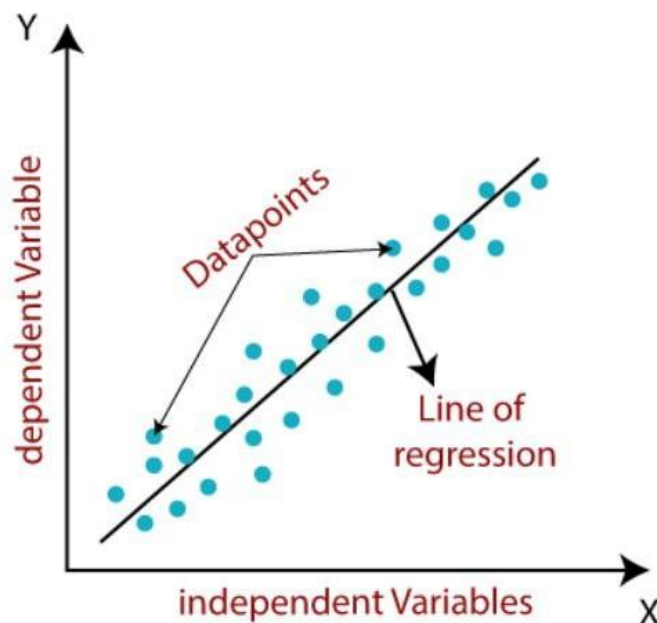
Y = Εξαρτώμενη μεταβλητή

X = Ανεξάρτητη μεταβλητή

a_0 = Τομή της γραμμής

a_1 = Συντελεστής γραμμικής παλινδρόμησης

ε = Σφάλμα



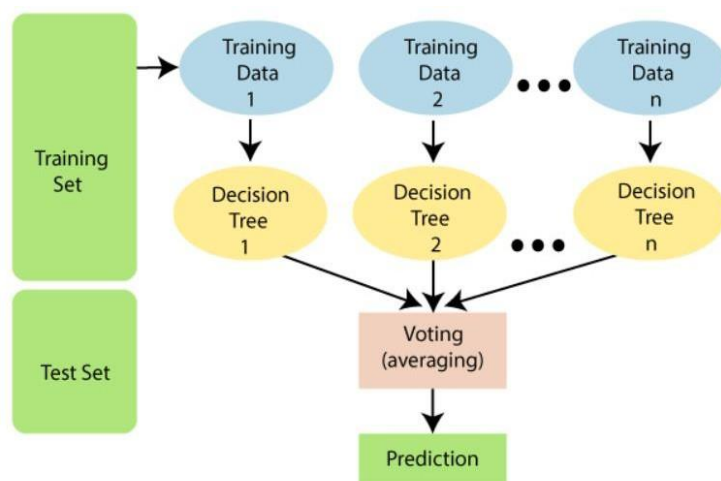
Εικόνα 8. Linear Regression

Random Forest

Ο Random Forest είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που ανήκει στην εποπτευόμενη τεχνική μάθησης. Μπορεί να χρησιμοποιηθεί για προβλήματα ταξινόμησης και παλινδρόμησης. Βασίζεται στην έννοια του ensemble learning, η οποία είναι μια διαδικασία συνδυασμού πολλαπλών ταξινομητών για την επίλυση ενός σύνθετου προβλήματος και τη βελτίωση της απόδοσης του μοντέλου.

Όπως υποδηλώνει το όνομα, ο Random Forest είναι ένας ταξινομητής που περιέχει έναν αριθμό δέντρων αποφάσεων σε διάφορα υποσύνολα του δοθέντος συνόλου δεδομένων και παίρνει τον μέσο όρο για να βελτιώσει την προγνωστική ακρίβεια αυτού του συνόλου δεδομένων. Αντί να βασίζεται σε ένα δέντρο αποφάσεων, ο αλγόριθμος λαμβάνει την πρόβλεψη από κάθε δέντρο και βασίζεται στις πλειοψηφίες των προβλέψεων και προβλέπει την τελική απόδοση.

Ο μεγαλύτερος αριθμός δέντρων στο δάσος οδηγεί σε μεγαλύτερη ακρίβεια και αποτρέπει το πρόβλημα της υπερφόρτωσης.



Εικόνα 9. Random Forest

2.3 Μάθηση Χωρίς Επίβλεψη

2.3.1 Συσταδοποίηση

K- means

Πρόκειται για ένα από τους πιο γνωστούς αλγόριθμους ομαδοποίησης. Υλοποιείται στον Ευκλείδειο χώρο και δέχεται ως όρισμα ένα προκαθορισμένο αριθμό συστάδων για να τμηματοποιήσει το σύνολο των δεδομένων σε ομάδες και να αναθέσει τα σημεία σε αυτές τις ομάδες. Ο αλγόριθμος στην πράξη προσπαθεί να ελαχιστοποιήσει μια μετρική που ονομάζεται Άθροισμα Τετραγωνικών Σφαλμάτων (SSE) και έτσι επιτυγχάνει την τμηματοποίηση του συνόλου δεδομένων σε συστάδες από σημεία με ασυμπτωτικά ίση διασπορά. Το SSE υπολογίζεται για κάθε σημείο της συστάδας και του κέντρου της συστάδας.

$$SSE = \sum_{x_i \in S_i} \min_{\mu_j \in C} (\|x_i - \mu_j\|^2) \quad (1)$$

Όπου μ_i είναι ο αριθμητικός μέσος όρος των σημείων στην συστάδα S_i και x_i το σημείο που πρόκειται να κατατάξουμε. Έτσι αν πάρουμε τον αλγόριθμο για το σύνολο των k συστάδων που έχουν δοθεί ως παράμετρο μπορούμε να εκφράσουμε το πρόβλημα με την εξίσωση:

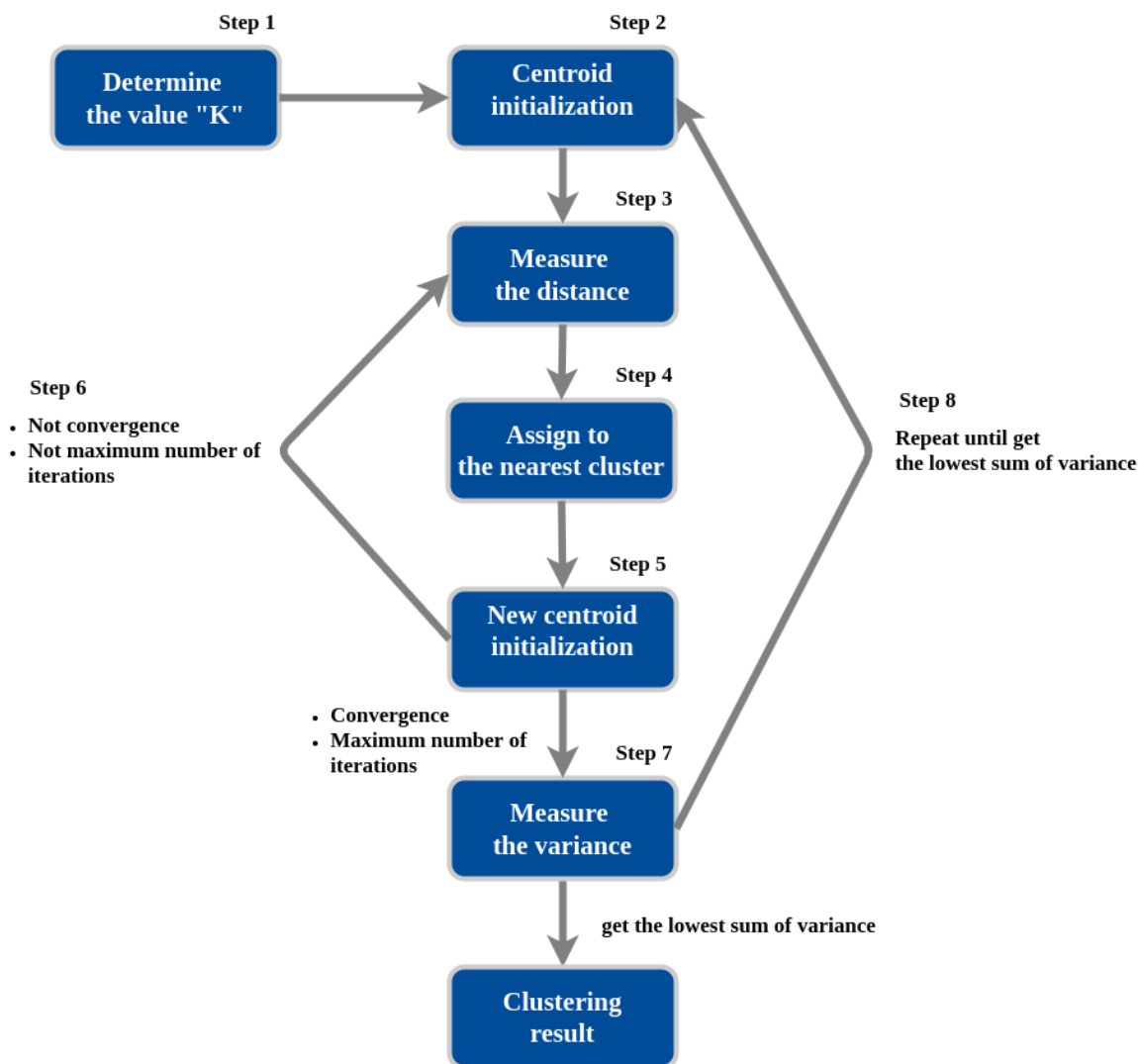
Όπως θα δούμε στην συνέχεια στο πειραματικό κομμάτι, από τις διάφορες εκδοχές του αλγορίθμου, έχει χρησιμοποιηθεί η υλοποίηση γνωστή ως Lloyd's algorithm ή αφελής μέθοδος. Σε αυτή την υλοποίηση ο αλγόριθμος εναλλάσσεται μεταξύ δύο καταστάσεων σε κάθε επανάληψη.

Βήμα 1. Ανάθεσε κάθε σημείο στην συστάδα με την μικρότερη Ευκλείδεια απόσταση από αυτό

$$S_i^{(t)} = \{ x_p : \|x_p - \mu_i\|^2 \leq \|x_p - \mu_j\|^2 \forall j, 1 \leq j \leq k \} \quad (2)$$

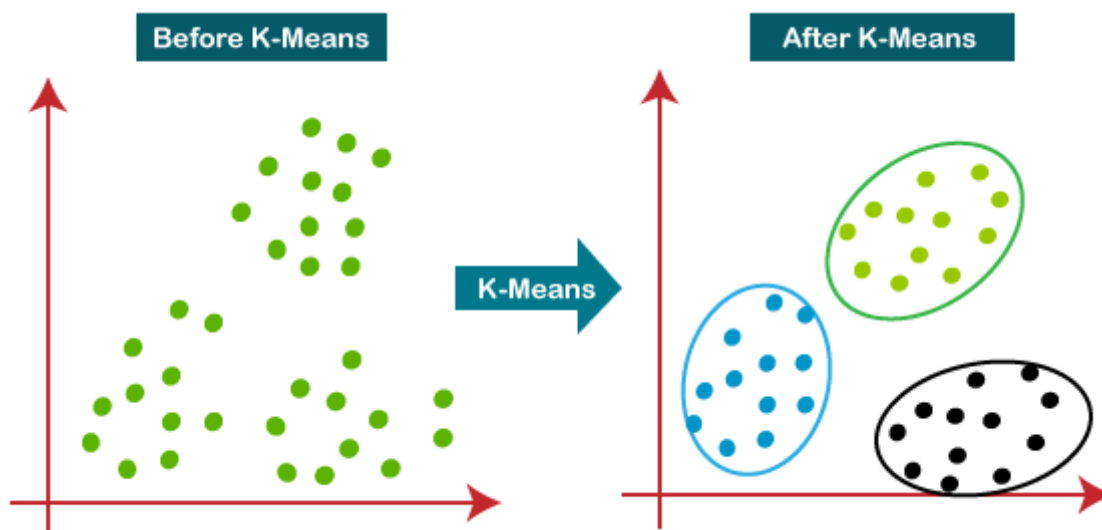
Βήμα 2. Υπολόγισε ξανά τα νέα κέντρα (τον αριθμητικό μέσο όρο των σημείων που ανήκουν στην συστάδα).

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_i \in S_i^{(t)}} x_i \quad (3)$$



Εικόνα 10. Αλγόριθμος K-means

Ο αλγόριθμος συγκλίνει όταν οι αναθέσεις των σημείων δεν αλλάζουν πλέον. Ο αλγόριθμος δεν εγγυάται την εύρεση της βέλτιστης λύσης. Στην φυσική σημασία του μπορούμε να θεωρήσουμε ότι ο αλγόριθμος επιλέγει ένα τυχαίο σύνολο από κέντρα για τις συστάδες, και αναθέτει σημεία στην συστάδα όπου το κέντρο της έχει την μικρότερη απόσταση από αυτό. Σε κάθε επανάληψη επαναυπολογίζεται το κέντρο κάθε συστάδας με τις εξισώσεις που παρουσιάσαμε (3). Χρησιμοποιώντας διαφορετικές μετρικές υπολογισμού της απόστασης από την Ευκλείδεια μπορεί να αποτρέψει τον αλγόριθμο από το να συγκλίνει ή και να οδηγήσει σε διαφορετικά αποτελέσματα.



Εικόνα 11. Παραγόμενες Συστάδες

Η συγκεκριμένη υλοποίηση που χρησιμοποιείται στην διπλωματική έχει μέση πολυπλοκότητα $O(k \cdot n \cdot T)$, όπου n ο αριθμός των σημείων, k ο αριθμός των συστάδων και T ο αριθμός των επαναλήψεων μέχρι να φτάσουμε σε αποτέλεσμα. Στην χειρότερη περίπτωση ο αλγόριθμος έχει πολυπλοκότητα $O(n^{(k + \frac{2}{p})})$

Στα πλεονεκτήματα είναι ότι ο αλγόριθμος είναι πολύ γρήγορος καθώς στην μέση περίπτωση έχουμε γραμμική πολυπλοκότητα (εκτός αν θέσουμε $k = n$, μια συστάδα για κάθε σημείο), ενώ στα αρνητικά είναι ότι δεν υπάρχει ντετερμινιστικός τρόπος εύρεσης του k , έχει ευαισθησία σε θόρυβο και outliers, και έχει κακά αποτελέσματα όταν οι συστάδες είναι διαφορετικού μεγέθους και πυκνότητας ή δεν έχουν σφαιρικό σχήμα, κάτι που είναι αρκετά συχνό στην δική μας περίπτωση όπου οι συστάδες αναπαριστούν σταθμούς βάσης και έχουμε σταθμούς βάσης ποικίλης εμβέλειας (Smallcells τύπου Microcells, Picocells, Femtocells)

PAM(K-medoids)

Ο K-medoids είναι παρόμοιος με τον K-means καθώς είναι και οι δύο αλγόριθμοι συσταδοποίησης που βασίζονται στην τμηματοποίηση, δηλαδή σπάζουν το σύνολο των δεδομένων σε ομάδες, και αποσκοπούν στην ελαχιστοποίηση των αποστάσεων κάθε σημείου από το κέντρο της συστάδας που του ανατέθηκε.

Σε αντίθεση με τον K-means, ο K-medoids διαλέγει ένα πραγματικό σημείο από το σύνολο των δεδομένων για να αναθέσει ως κέντρο, επιλέγουμε αυτό το σημείο που βρίσκεται στην ελάχιστη απόσταση από όλα τα υπόλοιπα ή που η μέση διαφορά του από όλα τα υπόλοιπα σημεία είναι η ελάχιστη. Από την άλλη ο K-means ανάθετε το κέντρο υπολογίζοντας την μέση τιμή των σημείων που βρίσκονται στην ομάδα. Επίσης ο K-medoids μπορεί να χρησιμοποιηθεί με αυθαίρετες μετρικές ομοιότητας σε αντίθεση με τον K-means που χρησιμοποιεί Ευκλείδειες αποστάσεις στην βασική του υλοποίηση. Με αυτό τον τρόπο ο PAM επιτυγχάνει να είναι πιο ανθεκτικός στον θόρυβο και σε outliers.

Υπάρχουν διάφοροι αλγόριθμοι που να υπολογίζουν την έννοια του K-medoids. Μερικοί από αυτού είναι:

- PAM (Partitioning around medoids)
- Clara (Clustering Large Applications)
- CLARANS (“Randomized” CLARA)

Από αυτούς ο PAM θεωρείται ο πιο ισχυρός και είναι ο πιο ευρέως χρησιμοποιημένος. Ωστόσο το μεγάλο ελάττωμα του PAM εμφανίζεται στην πολυπλοκότητα χρόνου.

Αλγόριθμος PAM(K-medoids)

Βήμα 1. Αρχικοποίησε k συστάδες από ένα δοσμένο σύνολο δεδομένων D διαλέγοντας k τυχαία σημεία ως τα κέντρα των συστάδων

Βήμα 2. Ανάθεσε κάθε σημείο στην κοντινότερη συστάδα χρησιμοποιώντας κάποια μετρική ομοιότητας. Στην περίπτωση μας καθώς έχουμε να κάνουμε με συντεταγμένες πραγματικού κόσμου (latitude, longitude) επιλέγουμε να χρησιμοποιήσουμε την μετρική **haversine** που είναι μια τριγωνομετρική συνάρτηση και χρησιμοποιείται για υπολογισμό αποστάσεων σε τρισδιάστατο επίπεδο

Βήμα 3. Όσο το συνολικό κόστος από την μετρική συνεχίζει να μειώνεται τότε:

Για κάθε σημείο p και κάθε κέντρο m

i) Αντικατέστησε το κέντρο m και το σημείο p μετατρέποντας το σημείο p ως το νέο κέντρο της συστάδας. Ανάθεσε κάθε σημείο που δεν είναι κέντρο στο κοντινότερο κέντρο δημιουργώντας μια νέα συστάδα και υπολογίζοντας τα νέα κόστη.

ii) Αν το συνολικό κόστος είναι μεγαλύτερο από πριν τότε αντέστρεψε την αντικατάσταση

Ο K-medoids λόγω της χρήσης medoid(κεντρικών σημείων) αντί της μέσης τιμής για την εύρεση κέντρου για την συστάδα μπορεί να θεωρηθεί και πιο εύρωστος σε σχέση με τον K-means. Αυτό συμβαίνει επειδή η μέση τιμή επηρεάζεται σε μεγαλύτερο βαθμό από την παρουσία θορύβου και ακραίων τιμών. Ωστόσο, όπως αναφέραμε το μεγάλο αρνητικό του K-medoids είναι η χρονική πολυπλοκότητα του, όπου η πολυπλοκότητα κάθε επανάληψης είναι $O(k(n-k)^2)$, για μεγάλες τιμές του n και του k οδηγούμαστε σε μεγάλους υπολογιστικούς χρόνους καθώς έχουμε να κάνουμε με πολυωνυμική πολυπλοκότητα.

Birch:

Οι προηγούμενες μέθοδοι που παρουσιάσαμε αν και αρκετά αποτελεσματικές με εύκολη παραμετροποίηση αδυνατούν στην διαχείριση μεγάλου όγκου δεδομένων όταν έχουμε να κάνουμε με περιορισμένο αριθμό πόρων (μνήμη, υπολογιστική ισχύ κλπ.). Έτσι καθώς ο όγκος των δεδομένων αυξάνεται τόσο χειρότερη και η κλιμάκωση τους σε ακρίβεια και χρόνο (ειδικά στην περίπτωση του αλγόριθμου PAM) αποτελεσμάτων. Ο αλγόριθμος **BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)**, όπου **εμπίπτει στην κατηγορία των αλγορίθμων ιεραρχικής συσταδοποίησης**, μπορεί να διαχειρίζεται μεγάλου όγκου δεδομένα δημιουργώντας μια πιο συμπαγές μορφή τους ενώ ταυτόχρονα διατηρεί την κατανομή της πληροφορίας, έτσι στην συνέχεια η συσταδοποίηση εφαρμόζεται στην συμπιεσμένη πληροφορία αντί στο αρχικό σύνολο δεδομένων. Αυτό επιτυγχάνεται αναπαριστώντας τα δεδομένα σε μια δενδρική δομή όπου με το πέρας της εκτέλεσης τα κέντρα βρίσκονται στα φύλλα.

Αυτή η συμπιεσμένη πληροφορία είναι γνωστή ως Clustering Feature (CF) και επιτρέπει την ελαχιστοποίηση των αναγκών σε μνήμη. Τα CF ορίζονται ως ένα σύνολο τριών τιμών, $CF = (N, LS, SS)$. Όπου:

$N =$ Πλήθος των σημείων στην Συστάδα

$$LS = \sum_{i=0}^{N-1} x$$

$LS =$ Το άθροισμα των N σημείων που ανήκουν στην Συστάδα

$$SS = \sum_{i=0}^{N-1} x^2$$

$SS =$ Το άθροισμα των τετραγώνων κάθε σημείου που ανήκει στην Συστάδα

Ουσιαστικά μπορούμε να θεωρήσουμε ότι κάθε CF αποτελεί η περιγραφή μιας ομάδας σημείων από το σύνολο δεδομένων που χειριζόμαστε. Χρησιμοποιώντας τα CF μπορούμε εύκολα να καταλήξουμε σε διάφορα στατιστικά στοιχεία για μία συστάδα:

Κέντρο
$$X_0 = \frac{\sum_{i=0}^{N-1} x_i}{n} = \frac{LS}{n}$$

Ακτίνα
$$\sqrt{\frac{\sum_{i=1}^n (x_i - x_0)^2}{n}} = \sqrt{\frac{nSS - 2LS^2 + nLS}{n^2}}$$

$$\text{Διάμετρος} \quad \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - x_o)^2}{n(n-1)}} = \sqrt{\frac{nSS - 2LS^2}{n(n-1)}}$$

Όπου η **ακτίνα** μας δίνει την μέση απόσταση ενός σημείου από το κέντρο της συστάδας στην οποία βρίσκεται και **διάμετρος** η μέση απόσταση από ένας ζεύγος σημείων μέσα σε μία συστάδα. Χρησιμοποιώντας αυτό το είδος συμπαγής πληροφορίας μας αποτρέπει από το να αποθηκεύουμε πληροφορίες για κάθε σημείο ξεχωριστά, και εδώ είναι που εμφανίζεται η ικανότητα του BIRCH να εξοικονομεί χώρο.

Μια χρήσιμη ιδιότητα των CF είναι ότι μπορούμε να τα συνδυάσουμε για να παράξουμε ένα νέο CF.

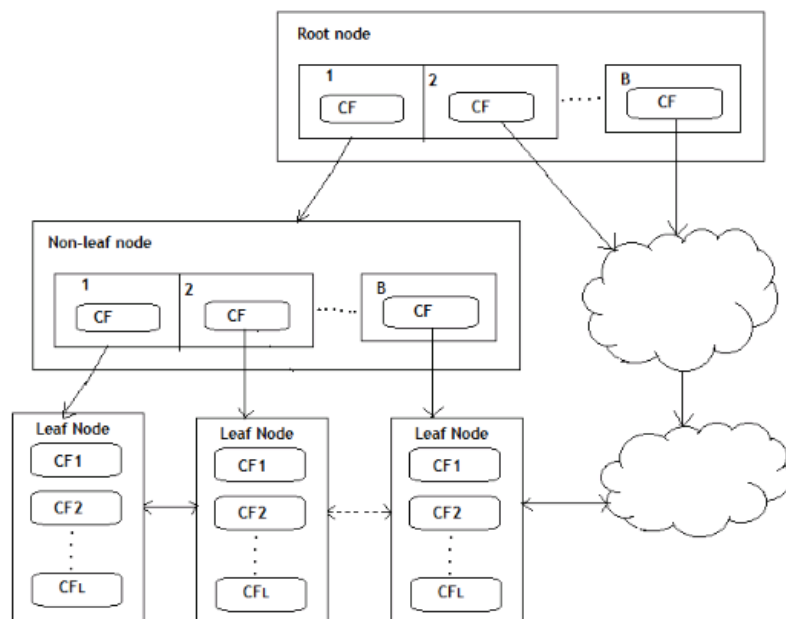
$$CF_1 + CF_2 = (N_1 + N_2, LS_1 + LS_2, SS_1 + SS_2)$$

Ο αλγόριθμος στον πυρήνα του στηρίζεται στην δημιουργία ενός CF-Tree. Αποτελεί μια ισοζυγισμένη συμπαγές δομή δεδομένων με 2 παράγοντες, τον παράγοντα διακλάδωσης B , που καθορίζει τον μέγιστο αριθμό παιδιών για κάθε κόμβο που δεν είναι φύλλο, και το κατώφλι T , που καθορίζει την μέγιστη διάμετρο υπο-συστάδων που μπορεί να έχει ένας κόμβος φύλλο [2]. Κάθε κόμβος που δεν είναι φύλλο πρόκειται για μια υπο-συστάδα που περιέχει μέχρι B καταχωρήσεις της μορφής $[CF_i, child_i]$, δηλαδή ένα ζεύγος από τα αθροισμένα CF που είναι παιδιά του i -οστού κόμβου και ένα δείκτη προς το i -οστό παιδί του κόμβου. Από την άλλη κάθε κόμβος φύλλο περιέχει μέχρι L καταχωρήσεις όπου κάθε μία είναι ένα CF (υπο-συστάδα σημείων). Κάθε καταχώρηση στα φύλλα πρέπει να πληρεί την προϋπόθεση ότι η διάμετρος του είναι μικρότερη ή ίση από το κατώφλι (T). Επίσης όπως φαίνεται στην Εικόνα 5, κάθε φύλλο περιέχει ένα δείκτη προς τον επόμενο και προηγούμενο κόμβο φύλλο.

Αλγόριθμος Κατασκευής CF-Tree

1. Για κάθε εγγραφή ο BIRCH συγκρίνει την εγγραφή με την θέση κάθε CF στην ρίζα, χρησιμοποιώντας το LS ή την μέση τιμή του CF. Στην συνέχεια παραχωρεί την νέα εγγραφή στον κόμβο ρίζα που βρίσκεται πλησιέστερα σε αυτήν
2. Η εγγραφή προσπελαύνει το υπο-δέντρο του κόμβου ρίζα που επιλέχθηκε στο Βήμα 1. Ο αλγόριθμος συγκρίνει την τοποθεσία κάθε κόμβου που δεν είναι φύλλο με την τοποθεσία της εγγραφής. Η εγγραφή παραχωρείται στον πλησιέστερο κόμβο που δεν είναι φύλλο και ανήκει στο υπο-δέντρο.
3. Η εγγραφή φτάνει στους κόμβους φύλλα του δέντρου που ανήκουν στον κόμβο που επιλέχθηκε στο Βήμα 2. Ο αλγόριθμος συγκρίνει την τοποθεσία της εγγραφής με τα παιδιά φύλλα του κόμβου. Μέσω επαναλήψεων ο αλγόριθμος αναθέτει την εγγραφή στον κόμβο φύλλο που βρίσκεται στην μικρότερη απόσταση.
Αν με την εισαγωγή της νέας εγγραφής η διάμετρος του κόμβου ξεπεράσει το Threshold τότε ο κόμβος διαιρείται

4. Επανάλαβε το βήμα (1) και (2).



Εικόνα 12. CF-Tree

Στην πιο πάνω εικόνα μπορούμε να δούμε πως κάθε κόμβος που δεν είναι φύλλο περιέχει B καταχωρήσεις, ενώ στα φύλλα έχουμε L Cluster Features (CF) και η διάμετρος τους δεν πρέπει να ξεπερνάει το Threshold. Έτσι μπορούμε να δούμε ότι κάθε φύλλο αποτελεί μια υπο-συστάδα.

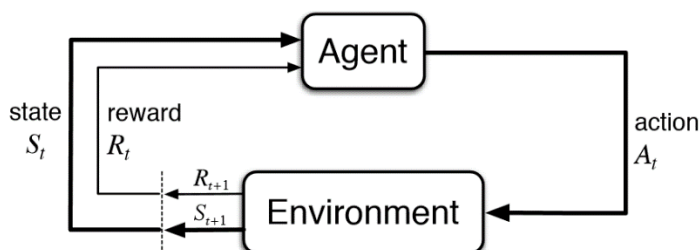
Η χρονική πολυπλοκότητα του **BIRCH** είναι $O(n)$, δηλαδή γραμμική, και όπου n ο αριθμός των αντικειμένων προς συσταδοποίηση. Μέσω πειραμάτων έχει αποδειχθεί η γραμμική κλιμάκωση του αλγορίθμου σε σχέση με τον αριθμό των αντικειμένων και η καλή ποιότητα συσταδοποίησης. Εντούτοις, δεδομένου ότι σε ένα δέντρο CF είναι περιορισμένος ο αριθμός εγγραφών που μπορεί να κρατήσει, λόγω των παραμέτρων B και T , ένας κόμβος του δέντρου δεν αντιστοιχεί πάντα σε αυτό που ένας χρήστης μπορεί να θεωρήσει πραγματική συστάδα. Επίσης, αν οι συστάδες δεν έχουν σφαιρικό σχήμα τότε ο **BIRCH** μπορεί να μην έχει την αναμενόμενη απόδοση καθώς χρησιμοποιεί την έννοια της ακτίνας και διαμέτρου για να ελέγξει τα όρια μιας συστάδας, αυτός είναι και ένας από τους λόγους που επιλέξαμε να δώσουμε στο πειραματικό κομμάτι της διπλωματικής σφαιρικό σχήμα στην αναπαράσταση της εμβέλειας ενός **Σταθμού Βάσης**[34].

2.4 Ενισχυτική Μάθηση

Με τον όρο Ενισχυτική Μάθηση (Reinforcement Learning-RL) κάνουμε λόγο για την κατηγορία της Μηχανικής Μάθησης που επιτρέπει σε ένα πράκτορα να μάθει μέσω της αλληλεπίδρασης του από το περιβάλλον, αυτό γίνεται μέσω της ανατροφοδότησης που παίρνει δοκιμάζοντας διάφορες ενέργειες όπου καταλήγουν σε αμοιβή ή αποτυχία.

Αν και η Εποπτευόμενη μάθηση χρησιμοποιεί αντιστοίχιση μεταξύ εισόδου και εξόδου, σε αντίθεση με την εποπτευόμενη μάθηση όπου η ανατροφοδότηση που παρέχεται στον πράκτορα είναι το σωστό σύνολο ενεργειών για την εκτέλεση μιας εργασίας, η ενισχυτική μάθηση χρησιμοποιεί ανταμοιβές ως ενδείξεις ότι οι ενέργειες που παίρνει το μοντέλο οδηγεί σε θετικό ή αρνητικό αποτέλεσμα.

Σε σύγκριση με τη μάθηση χωρίς επίβλεψη, η ενισχυτική μάθηση είναι διαφορετική ως προς τους στόχους. Ενώ ο στόχος στην μάθηση χωρίς επίβλεψη είναι να βρεθούν ομοιότητες και διαφορές μεταξύ των εισόδων, στην περίπτωση της ενισχυτικής μάθησης ο στόχος είναι να βρεθεί ένα κατάλληλο μοντέλο που θα μεγιστοποιήσει τη συνολική αθροιστική ανταμοιβή του πράκτορα.



Εικόνα 13 Βρόχος Ανταμοιβής

Σε ένα βασικό πρόβλημα Ενισχυτικής Μάθησης μοντελοποιημένο ως Markov Decision Process (MDP) χρειάζεται να ορίσουμε μερικά βασικά στοιχεία:

Πράκτορα: Μια οντότητα που δρα σε ένα περιβάλλον με σκοπό να μεγιστοποιήσει ένα μακροπρόθεσμο στόχο.

Περιβάλλον: Το σενάριο μέσα στο οποίο ο Πράκτορας λαμβάνει αποφάσεις.

Σύνολο Καταστάσεων (S): Το σύνολο όλων των πιθανών καταστάσεων του Περιβάλλοντος που περιγράφουν την παρούσα κατάσταση του Περιβάλλοντος.

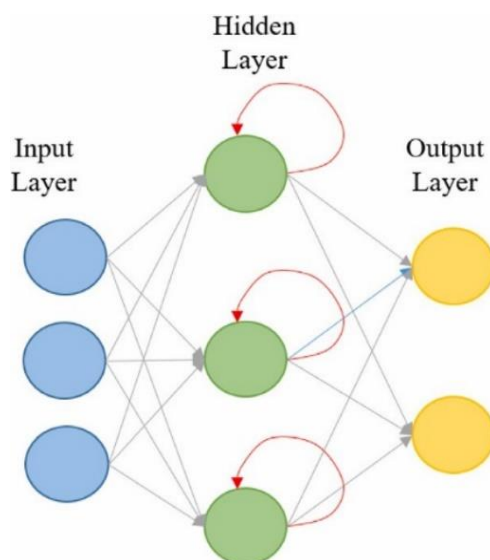
Σύνολο Ενεργειών (A): Το σύνολο όλων των πιθανών ενεργειών που μπορεί να πάρει ο Πράκτορας.

$P_a(s, s') = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a)$: Η πιθανότητα να αλλάξει η κατάσταση του περιβάλλοντος από $s \rightarrow s'$ όταν ο Πράκτορας εκτελέσει την ενέργεια a στην κατάσταση s την χρονική στιγμή t .

$R_a(s, s')$: Είναι η άμεση ανταμοιβή για την ενέργεια a που οδηγεί από την κατάσταση s στην κατάσταση s' [35].

2.5 Βαθιά Μάθηση (Deep Learning)

Αναδρομικά Νευρωνικά δίκτυα (RNN): Η αρχή του RNN είναι να αποθηκεύει την προηγούμενη έξοδο και να την τροφοδοτεί πίσω στην είσοδο ενώ έχει κρυφές καταστάσεις για να βοηθά τον αλγόριθμο στην πρόβλεψη του αποτελέσματος του επιπέδου. Για να το καταφέρει αυτό, το Νευρωνικό Δίκτυο χρησιμοποιεί διανύσματα κατάστασης (state vectors) που εξυπηρετούν τον ρόλο της μνήμης για τα εσωτερικά στοιχεία του NN που είναι οι νευρώνες. Επιπλέον, υπάρχουν συνδέσεις στα κρυφά επίπεδα της αρχιτεκτονικής RNN, όπου όλες οι εισοδοί, συμπεριλαμβανομένων των τρεχουσών και των πρώην εισόδων, επηρεάζουν την έξοδο. Ως εκ τούτου, ένα RNN έχει την ικανότητα να θυμάται. Το páρακατω σχήμα δείχνει τα κρυφά επίπεδα και τις συνδέσεις της αρχιτεκτονικής RNN.[5]



Εικόνα 14. Νευρωνικό Δίκτυο

LSTM

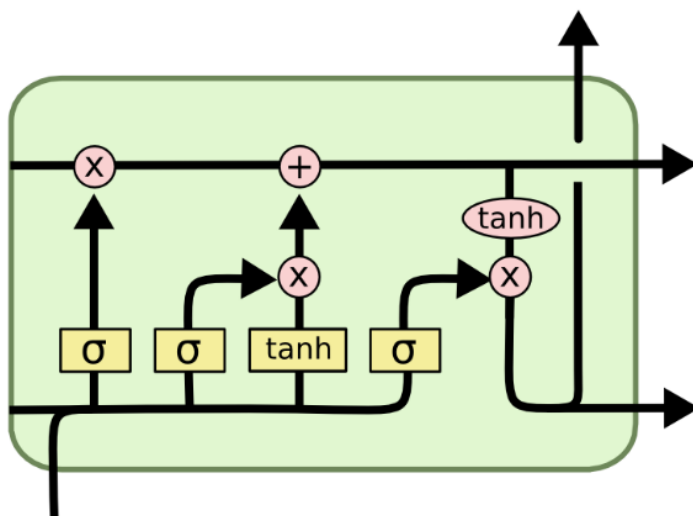
Τα Δίκτυα Μακροβραχυπρόθεσμης Μνήμης (LSTM) είναι μια αρχιτεκτονική τεχνητού επαναλαμβανόμενου νευρωνικού δικτύου (RNN) που χρησιμοποιείται στον τομέα της βαθιάς μάθησης. Σε αντίθεση με τα τυπικά νευρωνικά δίκτυα, το LSTM έχει συνδέσεις ανάδρασης. Μπορεί να επεξεργαστεί όχι μόνο μεμονωμένα σημεία δεδομένων (όπως εικόνες), αλλά και ολόκληρες ακολουθίες δεδομένων (όπως ομιλία ή βίντεο).

Η βασική έννοια των LSTM είναι η κυτταρική κατάσταση και οι διάφορες πύλες της. Η κατάσταση κυψέλης λειτουργεί ως αυτοκινητόδρομος μεταφοράς που μεταφέρει σχετικές πληροφορίες σε όλη την κατεύθυνση της αλυσίδας ακολουθιών, λειτουργεί δηλαδή σαν μια "μνήμη" του δικτύου. Η κυτταρική κατάσταση, θεωρητικά, μπορεί να μεταφέρει σχετικές πληροφορίες καθ' όλη τη διάρκεια της επεξεργασίας της αλληλουχίας. Έτσι, ακόμη και οι πληροφορίες από τα προηγούμενα χρονικά βήματα μπορούν να οδηγήσουν σε μεταγενέστερα χρονικά βήματα, μειώνοντας τις επιπτώσεις της βραχυπρόθεσμης μνήμης. Καθώς η κυτταρική κατάσταση συνεχίζει το ταξίδι της, οι πληροφορίες προστίθενται ή αφαιρούνται στην κατάσταση του κελιού μέσω πύλων. Οι πύλες είναι διαφορετικά νευρωνικά δίκτυα που

αποφασίζουν ποιες πληροφορίες επιτρέπονται για την κατάσταση του κυττάρου. Οι πύλες μπορούν να μάθουν ποιες πληροφορίες είναι σχετικές για να διατηρηθούν ή να ξεχαστούν κατά τη διάρκεια της εκπαίδευσης. Αυτή η τεχνολογία ήρθε στο αναπτύχθηκε με σκοπό να λύσει ένα πρόβλημα που αντιμετώπιζαν τα μέχρι τώρα Νευρωνικά Δίκτυα, το πρόβλημα της εξαφανιζόμενης κλίσης (vanishing gradient). [6]

Οι πύλες περιέχουν σιγμοειδείς ενεργοποιήσεις. Μια ενεργοποίηση σιγμοειδούς είναι παρόμοια με την ενεργοποίηση tanh, η οποία συμπιέζει τις τιμές μεταξύ 0 και 1. Αυτό είναι χρήσιμο για την ενημέρωση ή τη λήξη δεδομένων, επειδή οποιοσδήποτε αριθμός πολλαπλασιάζεται με το 0 είναι 0, προκαλώντας εξαφάνιση ή "λήθη" των τιμών. Οποιοσδήποτε αριθμός πολλαπλασιασμένος με 1 είναι η ίδια τιμή, επομένως η τιμή παραμένει ίδια ή "διατηρείται". Το δίκτυο μπορεί να μάθει ποια δεδομένα δεν είναι σημαντικά, επομένως μπορούν να ξεχαστούν ή ποια δεδομένα είναι σημαντικό να διατηρηθούν.

Σε ένα LSTM υπάρχουν 3 διαφορετικά είδη πυλών (πύλη λήψης, πύλη εισόδου, πύλη εξόδου). Η πύλη λήψης αποφασίζει ποιες πληροφορίες πρέπει να πετιούνται ή να φυλάσσονται. Η πύλη εισόδου χρησιμοποιείται για να ενημερώνει τη κατάσταση κελιού. Τέλος, η πύλη εξόδου αποφασίζει ποια πρέπει να είναι η επόμενη κρυφή κατάσταση, η οποία περιέχει πληροφορίες για προηγούμενες εισόδους και μπορεί να χρησιμοποιηθεί ακόμη και για προβλέψεις.



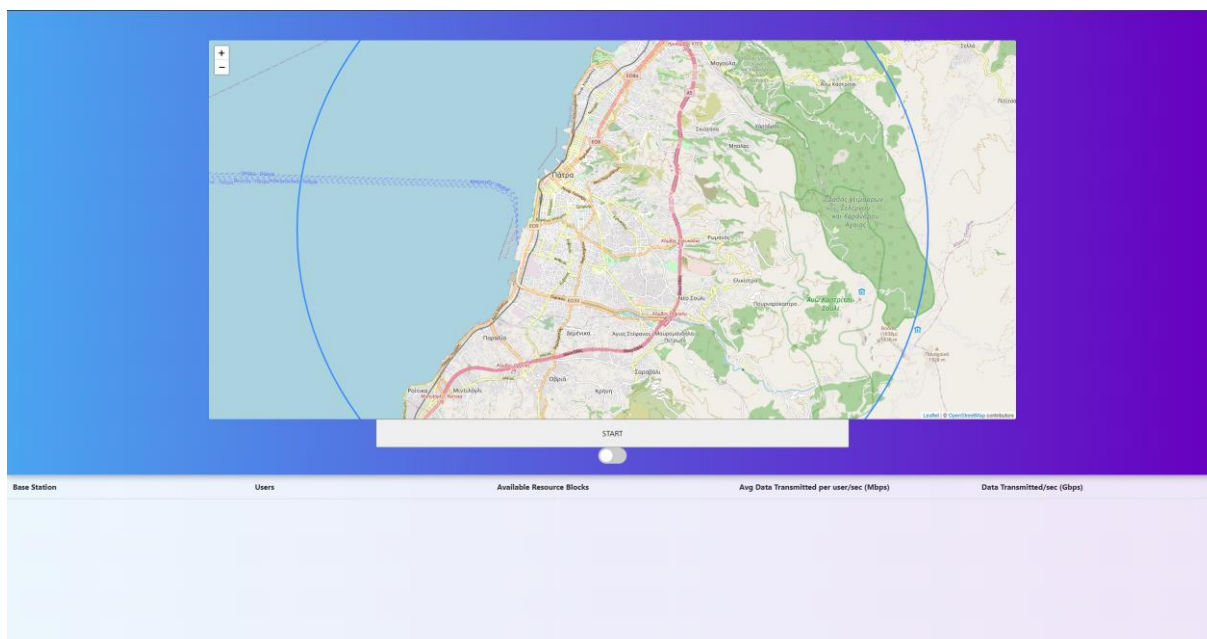
Εικόνα 15. NN LSTM

3

ΠΕΡΙΓΡΑΦΗ ΓΡΑΦΙΚΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

3.1 Δημιουργία Γραφικού Περιβάλλοντος

Για τα πλαίσια αυτής της διπλωματικής διενεργήθηκε η υλοποίηση μιας Διεπαφής Γραφικού Περιβάλλοντος (Graphical User Interface – GUI), που θα εξυπηρετήσει τον σκοπό του εργαλείου και μέσου για την απεικόνιση των διαφόρων μεθόδων και πρακτικών που θα παρουσιάσουμε στην συσχέτιση των UE σε Σταθμούς Βάσης. Η προσέγγιση που ακολουθήθηκε για την υλοποίηση του GUI ήταν η δημιουργία μιας διαδικτυακής εφαρμογής, υλοποιημένη χρησιμοποιώντας βασικά στοιχεία ανάπτυξης διαδικτυακών εφαρμογών όπως **HTML5**, **CSS**, **JavaScript** (jQuery) ως προς το κομμάτι του frontend, και που αποτελείται από μια σελίδα που απεικονίζει έναν χάρτη εστιασμένο στο κέντρο της πάτρα. Για την δημιουργία και διαχείριση του χάρτη χρησιμοποιήθηκε η βιβλιοθήκη **Leaflet**, μια βιβλιοθήκη της JavaScript για την απεικόνιση σημείων με πραγματικές συντεταγμένες αντί καρτεσιανών. Η σελίδα πριν την έναρξη του πειράματος φαίνεται παρακάτω.



Εικόνα 16. Γραφικό Περιβάλλον

Ξεκινώντας από πάνω προς τα κάτω βλέπουμε τον χάρτη όπου θα γίνει η απεικόνιση όλων των στοιχείων που λαμβάνουν μέρος στην οργάνωση ενός δικτύου 5G. Στην συνέχεια βλέπουμε ένα στοιχείο κουμπιού όπου με το πάτημα του ξεκινά η προσομείωση (οποιαδήποτε στιγμή μπορεί να γίνει παύση της προσομοίωσης), ένα στοιχείο διακόπτη όπου με την ενεργοποίηση του επιτρέπουμε στον χρήστη την εισαγωγή νέων σημείων (UE) στην προσομείωση πέρα από αυτών που παράγονται από την ρύθμιση των παραμέτρων της προσομοίωσης. Τέλος βλέπουμε τον πίνακα όπου κατά την διάρκεια της προσομοίωσης απεικονίζονται διάφορα στατιστικά των Σταθμών Βάσης που θα παραχθούν με δυναμικό τρόπο από τον προσομοιωτή. Αυτά τα στατιστικά στοιχεία είναι:

- Το ID κάθε Σταθμού Βάσης
- Ο Αριθμός των χρηστών που ανατέθηκαν σε κάθε Σταθμό Βάσης
- Οι διαθέσιμοι πόροι (Resource Blocks) κάθε Σταθμού Βάσης
- Ο μέσος ρυθμός μετάδοσης δεδομένων προς τα συνδεδεμένα UE κάθε Σταθμού Βάσης σε Mbps
- Ο συνολικός όγκος δεδομένων που μεταφέρεται από τον Σταθμό Βάσης προς τα συνδεδεμένα UE σε Gbps

Η μεριά του Backend έχει υλοποιηθεί χρησιμοποιώντας python3, ενώ όλες τις ενέργειες του server τις διαχειριζόμαστε χρησιμοποιώντας την βιβλιοθήκη **flask**. Χρησιμοποιώντας το **flask** ακούμε για http requests και επιστρέφουμε τα κατάλληλα δεδομένα, στην περίπτωση μας είναι το σύνολο των UE και BS προς απεικόνιση συνοδευόμενα με τα απαραίτητα χαρακτηριστικά.

Στο Backend οι Χρήστες και οι Σταθμοί Βάσης αναπαριστούντε ως κλάσεις. Κάθε κλάση περιέχει τις αναγκαίες μεταβλητές για την εκτέλεση της προσομοίωσης.

```

class Point:
    var = 0
    distance_block = 500/28
    # resource_demands = {'picture': 50, 'video': 1100, 'gaming': 650, 'streaming': 750, 'web browsing':250}

    def __init__(self, lat, lon, centerLat, centerLon):
        self.lat = lat
        self.lon = lon
        self.centerLat = centerLat
        self.centerLon = centerLon
        self.BSId = -1
        self.id = Point.var
        self.distance = calcDistance(lat, lon, centerLat, centerLon)
        self.BSarray = []
        self.throughput = 0
        # self.action = random.choice(list(Point.resource_demands.keys()))
        self.action = random.randint(50, 1500)
        self.resource_block_demands = 0
        self.SINR = 0
        Point.var += 1

    def fill_BSarray(self):...

    def calculateNoise(self):...

    def calculate_throughput(self, bandwidth, mino_layers):
        number_of_symbols = 11
        subcarriers = 12
        resource_elements = subcarriers * number_of_symbols
        index = 28 - int(self.distance * 1000 // Point.distance_block)
        self.SINR = round(random.uniform(SINR[index-1][0], SINR[index-1][1]), 2)
        bits_per_symbol = MCS[index - 1]
        data_per_resource_block = resource_elements * bits_per_symbol #bits
        max_number_of_RBs = bandwidth / (subcarriers*30) - 4
        Ts = (1 / (1000 * 14 * 2))
        throughput = (4 * MCS2[index - 1] * 1 * (codeRate[index - 1]/1024) * ((self.resource_block_demands * 12) / Ts) * (1 - 0.14)) * 4
        self.throughput = throughput / 1000000

```

Κώδικας 1. Κλάση Point

Η κλάση `Point` εκτελεί τον ρόλο ενός UE. Περιέχει μεταβλητές όπως:

- Οι μεταβλητές **lat**, **lon** που αποθηκεύουν το Latitude και Longitude (δηλαδή τις γεωγραφικές συντεταγμένες του χρήστη) και μεταβάλλονται με τυχαίο τρόπο σε κάθε επανάληψη καθώς θεωρούμε ότι οι χρήστες βρίσκονται σε συνεχόμενη κίνηση.
- Οι μεταβλητές **centerLat**, **centerLon** που είναι οι γεωγραφικές συντεταγμένες του κέντρου του BS από το οποίο εξυπηρετείται ο χρήστης σε κάθε επανάληψη. Αρχικοποιούνται στις τιμές [0.0,0.0] και παραμένουν σε αυτές τις τιμές αν ο χρήστης δεν εξυπηρετείται από κάποιο BS.
- Η μεταβλητή **BSid** που αποθηκεύει το μοναδικό αναγνωριστικό (ID) του BS από τον οποίο εξυπηρετείται ο χρήστης. Αν δεν εξυπηρετείται από κάποιο BS τότε παίρνει την τιμή -1.
- Το μοναδικό αναγνωριστικό του χρήστη (**id**).
- Την μεταβλητή **distance** που αποθηκεύει την απόσταση του χρήστη από το κέντρο του BS που τον εξυπηρετεί και υπολογίζεται μεταξύ των γεωγραφικών συντεταγμένων χρησιμοποιώντας την φόρμουλα Haversine.
- Την μεταβλητή **BSarray** που είναι μια λίστα που αποθηκεύει όλα τα στιγμιότυπα Σταθμών Βάσης στα οποία ο χρήστης βρίσκεται στην εμβέλεια τους. Αρχικά είναι κενή και με το κάλεσμα της συνάρτησης `fill_BSarray()` τότε διατρέχουμε όλα τα ενεργά στιγμιότυπα από BS που έχουμε δημιουργήσει και υπολογίζουμε αν η απόσταση του χρήστη είναι μικρότερη από την εμβέλεια του BS, τότε αυτό το BS προστίθεται στην λίστα.
- Η μεταβλητή **throughput** που αποθηκεύει τον ρυθμό μετάδοσης δεδομένων μεταξύ του Χρήστη και του BS που τον εξυπηρετεί. Υπολογίζεται καλώντας την συνάρτηση `calculate_throughput` που υλοποιεί την εξίσωση (5). Θεωρούμε για τους σκοπούς αυτή της προσομοίωσης ότι έχουμε μόνο Downlink μετάδοση.
- Την μεταβλητή **action** που αποθηκεύει τις απαιτήσεις του χρήστη σε throughput (Mbps) και πρόκειται για ένα τυχαίο ακέραιο στο διάστημα [50,1500].

- Την μεταβλητή **resource_block_demands** που αποθηκεύει τις απαιτήσεις του χρήστη από τον Σταθμό Βάσης σε πόρους (Resource Blocks).
- Την μεταβλητή **SINR** που αποθηκεύει το SINR του χρήστη και υπολογίζεται με την εξίσωση (2).

```
class BaseStation:
    USER_THRESHOLD = 50
    var = 0
    bandwidth = 100 #MHz
    MAX_RESOURCE_BLOCK_NUMBER = 273 # 4 RBs are kept for guard, bandwidth / (subcarriers * scs frequency)

    def __init__(self, lat, lon):
        self.lat = lat
        self.lon = lon
        self.id = BaseStation.var
        self.users = []
        self.resource_blocks_available = BaseStation.MAX_RESOURCE_BLOCK_NUMBER
        self.average_throughput = 0
        BaseStation.var += 1

    def addUser(self, point):
        self.users.append(point)

    def check(self):
        if len(self.users) <= BaseStation.USER_THRESHOLD:
            return True
```

Κώδικας 2. Κλάση Σταθμού Βάσης (BS)

Η κλάση BS εκτελεί τον ρόλο ενός Σταθμού Βάσης. Περιέχει μεταβλητές όπως:

- Οι μεταβλητές **lat**, **lon** που αποθηκεύουν το Latitude και Longitude (δηλαδή τις γεωγραφικές συντεταγμένες του BS) και υπολογίζονται από τον εκάστοτε αλγόριθμο συσταδοποίησης που διαλέγουμε να χρησιμοποιήσουμε μαζί με τον αριθμό συστάδων που ορίζουμε. Δεν μεταβάλλονται.
- Το μοναδικό αναγνωριστικό του BS (**id**).
- Την μεταβλητή **users**, μια λίστα που αποθηκεύει τα στιγμιότυπα χρηστών που εξυπηρετεί ο Σταθμός Βάσης. Η προσθήκη γίνεται χρησιμοποιώντας την συνάρτηση addUser() ώστε να πετύχουμε ενθυλάκωση ως προς τα δεδομένα του στιγμιότυπου.

- Την μεταβλητή **resource_blocks_available** που αποθηκεύει τους διαθέσιμους πόρους του BS σε κάθε επανάληψη και αρχικοποιείται στον μέγιστο αριθμο από RBs που μπορεί αν περιέχει ο BS, για αυτό το παράδειγμα Macrocells. Στην συγκεκριμένη προσομοίωση διαλέγουμε το δίκτυο να έχει 100 MHz bandwidth, 12 subcarriers για κάθε Resource block με 30 KHz subcarrier spacing. Έτσι ο μέγιστος αριθμός από RBs προκύπτει από:

$$Max\ Number\ of\ RBs = \frac{bandwidth}{Number\ of\ subcarriers * subcarrier\ spacing} - 4 \quad (1)$$

4 RBs δεν είναι διαθέσιμα για χρήση καθώς κρατιούνται ως ασφάλεια. Με τις δοσμένες παραμέτρους του δικτύου προκύπτει ότι ο μέγιστος αριθμός από RBs για κάθε BS είναι 273. Ο διαθέσιμος αριθμός από RBs για κάθε BS είναι άμεσα εξαρτόμενος από την διαμόρφωση των subcarriers και διαφέρουν ανάλογα με το πως ρυθμίζονται.[9]

- Την μεταβλητή **average_throughput** που αποθηκεύει τον μέσο ρυθμό μετάδοσης δεδομένων του BS προς τους χρήστες.

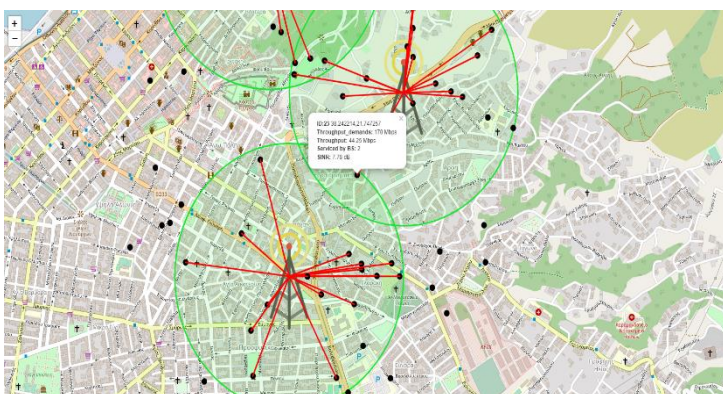
Δημιουργία Τυχαίων Σημείων (UE)

```
//RANDOM POINTS GENERATED
var points = [];
for (var j = 0; j < users; j++) {
  var r = gFactor * Math.sqrt(Math.random());
  var theta = Math.random() * 2 * Math.PI;
  var x = 38.24306756139534 + r * Math.cos(theta);
  var y = 21.74490490893847 + r * Math.sin(theta);
  points[j] = {'lat': x.toFixed( fractionDigits: 6), 'lon': y.toFixed( fractionDigits: 6), 'id': j, 'throughput': 0,
    'throughput_demands': 0, 'serviced': -1, 'SINR': 0};
}
```

Κώδικας 3. Δημιουργία Τυχαίων Σημείων

Για τον σκοπό της προσομοίωσης δημιουργούμε ένα αριθμό σημείων με τυχαίες συντεταγμένες που παριστάνουν τα UE. Χρησιμοποιώντας την μεταβλητή `gFactor` περιορίζουμε τον τυχαίο αριθμό που επιστρέφει η γεννήτρια τυχαίων αριθμών `Math.random()` με ομοιόμορφη κατανομή στο διάστημα $[0,1]$, έτσι ώστε να ελέγχουμε το πόσο ομαδοποιημένα θέλουμε να είναι τα σημεία. Δηλαδή μπορούμε να ελέξουμε την διασπορά των σημείων γύρω από το κεντρικό σημείο. Όσο μικρότερη η `gFactor` τόσο πιο ομαδοποιημένα θα είναι τα σημεία. Ως σημείο αναφοράς χρησιμοποιούμε τις συντεταγμένες της περιοχής της Πάτρας και προσθέτουμε σε αυτές κάθε φορά την τυχαία μετατόπιση. Έπειτα κάθε σημείο αποθηκεύεται ως ένα JSON object, με τα κατάλληλα πεδία, στον πίνακα `points`. Αυτά τα σημεία στην συνέχεια στέλνονται στον server για να εφαρμόσει την συσταδοποίηση και να μας επιστρέψει τις συντεταγμένες και τον αριθμό των κέντρων για κάθε Σταθμό Βάσης.

Απεικόνιση Σημείων



Εικόνα 17. Απεικόνιση σημείων στον Χάρτη

```
//CREATE THE POINT FOR EACH USER (THE BLACK DOT)
for (var j = 0; j < Object.keys(newData).length; j++) {
  L.circle([newData[j]['lat'], newData[j]['lon']], {
    fill: true,
    fillOpacity: 1,
    fillColor: '#000000',
    color: '#000000',
    radius: 10
  }).addTo(group);

  //CREATE POPUP FOR EACH USER SHOWING HIS ID AND COORDINATES
  group.getLayers().at(-1).bindPopup("ID: " + newData[j]['id'].toString().bold() + " "
    + [newData[j]['lat'], newData[j]['lon']].toString() + "<br>"
    + "Throughput: " + newData[j]['throughput'].toString() + " Mbps" + "<br>"
    + "Serviced by BS: " + newData[j]['serviced'].toString() + " Mbps" + "<br>"
    + "SINR: " + newData[j]['SINR'].toString() + " dB", {maxWidth: 500});

  var BSid = newData[j]['serviced'];

  // DIMITOURGOME TA KANALIA
  if (BSid != -1) {
    bs_line = centersBS[BSid];
    point_line = [newData[j]['lat'], newData[j]['lon']];
    L.polyline([bs_line, point_line], {color: 'red'}).addTo(group);
    line_tooltip[newData[j]['id']] = group.getLayers().at(-1);
  }
}
```

Κώδικας 4. Απεικόνιση Σημείων

Σε κάθε επανάληψη αφαιρούμε από τον χάρτη όλα τα προηγούμενα σημεία (εκτός αν πρόκειται για την πρώτη επανάληψη) και χρησιμοποιώντας την ιδιότητα της

βιβλιοθήκη leaflet circle δημιουργούμε ένα αντικείμενο τύπου circle με τις συντεταγμένες των τυχαίων σημείων και τα τοποθετούμε στον χάρτη. Αφαιρώντας κάθε προηγούμενα σημεία και προσθέτοντας τα ξανά (με μια σταθερή μετατόπιση μεγέθους 40 μέτρων και τυχαία επιλεγμένη κατεύθυνση από βορά, νότο, δύση, ανατολή) δημιουργούμε την ψευδαίσθηση της κίνησης. Σημαντικό είναι επίσης να αναφέρουμε ότι κάθε σημείο είναι μοναδικό και φαίνεται από το αναγνωριστικό του (ID), έτσι μπορούμε να αντιμετωπίζουμε κάθε σημείο ως μια πραγματική οντότητα και δεν χρειάζεται σε κάθε επανάληψη η δημιουργία νέων σημείων.

Για κάθε σημείο ορίζουμε ένα Popup με αποτυπωμένες τις παραμέτρους που περιέχουν τα UE στην προσομοίωση, πυροδοτείται πατώντας το σημείο όταν η προσομοίωση βρίσκεται σε παύση. Αυτές οι παράμετροι ορίζονται στην μεριά του server ανάλογα με την κατάσταση του σημείου και τα δεδομένα του δικτύου, έπειτα επιστρέφονται στο frontend κομμάτι της εφαρμογής. Στο Popup κάθε σημείου μπορούμε να διακρίνουμε τα εξής χαρακτηριστικά:

- Το μοναδικό του αναγνωριστικό που μας βοηθάει να το αντιμετωπίζουμε ως ένα πραγματικό UE μαζί με τις πραγματικές γεωγραφικές συντεταγμένες του.
- Τις απαιτήσεις του UE σε πόρους αποτυπωμένες σε Throughput
- Το Throughput που μπορεί να επιτευχθεί από τον Σταθμό Βάσης λαμβάνοντας υπόψη διάφορες παραμέτρους όπως SINR, διαθέσιμα RBs, CQI value και ως εκ τούτου αντίστοιχο Modulation Coding Scheme κτλπ.
- Τον Σταθμό Βάσης από τον οποίο εξυπηρετείται. Αν δεν έχει κάποιος Σταθμός Βάσης τα απαραίτητα RBs για να τον εξυπηρετήσει ή δεν βρίσκεται στην εμβέλεια κάποιου BS θα εμφανίζεται '-1'.
- Το SINR όπως υπολογίζεται από τον τύπο (2).

Για κάθε UE δημιουργούμε ένα κανάλι, ως μια κόκκινη γραμμή, προς τον BS που τον εξυπηρετεί χρησιμοποιώντας την κλάση της leaflet – polyline, όπου δίνουμε ως όρισμα τις συντεταγμένες του σημείου και του κέντρου του BS. Όπως και με τα σημεία αφαιρούμε κάθε προηγούμενες γραμμές πριν την τοποθέτηση των νέων.

Απεικόνιση Σταθμών Βάσης

```

//BASE STATION
for (var i = 0; i < Object.keys(data).length; i++) {
  L.imageOverlay("base-station.jpg", [data[i][1], data[i][2], data[i][3], data[i][4]]).addTo(map);
  circlesBS[i] = L.circle(data[i][0], {
    fill: false,
    radius: 500,
    color: '#0fe92c'
  }).addTo(map);

  circlesBS[i].bindPopup("No." + i.toString() + " " + data[i][0].toString(), {maxWidth: 500});
  // circlesBS[i].on('click', function () {
  //   circlesBS[i].openPopup(data[i][0]);
  // });

  //STORE EACH BSs COORDINATES IN GLOBAL VAR
  centersBS[i] = data[i][0];

  //CREATE THE HTML TABLE
  var row = table.insertRow( index: i + 1);
  row.insertCell( index: 0).innerHTML = "No. " + i.toString();
  row.insertCell( index: 1);
  row.insertCell( index: 2);
  row.insertCell( index: 3);
  row.insertCell( index: 4);
}
table.insertRow( index: i + 1).insertCell( index: 0).innerHTML = "Total";
table.rows[table.rows.length - 1].insertCell( index: 1);

```

Κώδικας 5. Απεικόνιση BS

Σε αντίθεση με την δημιουργία σημείων που αντιπροσωπεύουν UE, η διαδικασία δημιουργίας BS εκτελείται μόνο μια φορά κατά την δημιουργία του GUI και καθώς η προσομοίωση βρίσκεται ακόμη σε παύση. Στέλνοντας ένα ajax request στον server λαμβάνουμε πίσω τα κέντρα των Σταθμών Βάσης, αφού έχει προηγηθεί η συσταδοποίηση των σημείων. Προσπελάζουμε το response object και για κάθε κέντρο που δημιούργησε ο αλγόριθμος συσταδοποίησης δημιουργούμε δύο αντικείμενα, ένα τύπου circle που αντιπροσωπεύει την εμβέλεια του Σταθμού Βάσης (θεωρούμε ότι οι Σταθμοί Βάσης είναι Macrocells με εμβέλεια τα 500 μετρα και κάλυψη που επεκτείνεται στην μορφή κύκλου), και ένα αντικείμενο τύπου imageOverlay που είναι η εικόνα ενός Σταθμού Βάσης στο κέντρο του κύκλου. Σε κάθε BS επισυνάπτουμε ένα Popup όπου αναγράφονται οι συντεταγμένες του BS και το μοναδικό αναγνωριστικό του (ID). Τα υπόλοιπα χαρακτηριστικά για κάθε Σταθμό Βάσης είναι διαθέσιμα στον πίνακα που δημιουργούμε στο τέλος της σελίδας με δυναμικό τρόπο και περιέχει μια γραμμή για κάθε BS.

Base Station	Users	Available Resource Blocks	Avg Data Transmitted per user/sec (Mbps)	Data Transmitted/sec (Gbps)
No. 0	16	3	349.02	5.58
No. 1	17	8	268.76	4.57
No. 2	15	78	203.51	3.05
Total	48 / 80			

Εικόνα 18 Πίνακας Στοιχείων BS

Όπως φαίνεται για κάθε BS μπορούμε να δούμε σε κάθε επανάληψη τον αριθμό των χρηστών που εξυπηρετεί, τον αριθμό των διαθέσιμων RBs, το μέσο Throughput των χρηστών που εξυπηρετεί όπως υπολογίζεται από την εξίσωση (5), και τον συνολικό αριθμό όγκου δεδομένων που μεταφέρεται από τον Στθαμό Βάσης προς τα UE.

Πριν συνεχίσουμε στο σημαντικότερο κομμάτι αυτής της διπλωματικής, και τον πυρήνα του πειράματος μας, είναι σημαντικό να ορίσουμε μερικά σημαντικά ζητήματα που αφορούν τα δίκτυα 5ης Γενιάς, και ο τρόπος υπολογισμού τους αλλά και η χρησιμοποίησή τους επηρεάζουν την έκβαση των αποτελεσμάτων μας άμεσα.

Υπολογισμος SINR

Το SINR χρησιμοποιείται συνήθως στην ασύρματη επικοινωνία ως τρόπος μέτρησης της ποιότητας των ασύρματων συνδέσεων. Τυπικά, η ενέργεια ενός σήματος εξασθενεί με την απόσταση, η οποία αναφέρεται ως απώλεια διαδρομής (Path Loss – PL) στα ασύρματα δίκτυα. Αντίθετα, στα ενσύρματα δίκτυα η ύπαρξη μιας ενσύρματης διαδρομής μεταξύ του αποστολέα ή του πομπού και του δέκτη καθορίζει τη σωστή λήψη δεδομένων. Σε ένα ασύρματο δίκτυο πρέπει να ληφθούν υπόψη άλλοι παράγοντες (π.χ. ο θόρυβος του περιβάλλοντος, η ισχύς παρεμβολής άλλης ταυτόχρονης μετάδοσης). Η έννοια του SINR επιχειρεί να δημιουργήσει μια αναπαράσταση αυτής της πτυχής.[12][13]

Αποτελεί κοινή πρακτική η χρήση του SINR ως δείκτη για την ποιότητα του δικτύου. Θα πρέπει ωστόσο να σημειωθεί ότι οι προδιαγραφές 3GPP δεν ορίζουν το SINR και επομένως το UE (User Equipment) αναφέρει το SINR στο δίκτυο. Το SINR εξακολουθεί να μετράται εσωτερικά από τα περισσότερα UE και να καταγράφεται από τα εργαλεία δοκιμής κίνησης. Δυστυχώς, οι κατασκευαστές UE chipset και σαρωτών RF έχουν εφαρμόσει τη μέτρηση SINR με διάφορους τρόπους. Ενώ στην αρχή μπορεί να φαίνεται ότι ο ορισμός του SINR θα πρέπει να είναι ξεκάθαρος, στην περίπτωση Downlink σε 5G αυτό δεν ισχύει. Αυτό οφείλεται στο ότι διαφορετικά Resource Elements μέσα σε ένα frame μεταφέρουν διαφορετικά φυσικά σήματα και κανάλια, καθένα από τα οποία, με τη σειρά του, βλέπει διαφορετικά επίπεδα παρεμβολής ανάλογα με τον συγχρονισμό των πλαισίων μεταξύ των κυψελών. Η μέτρηση του SINR μπορεί να χρησιμοποιηθεί σε διάφορους τομείς στην ανάλυση δικτύων, όπως για τον υπολογισμό της τιμής CQI(Channel Quality Information).

Μια καλή τιμή SINR μπορεί να βοηθήσει στην επίτευξη υψηλότερης φασματικής απόδοσης, καθώς επιτρέπει την αποκωδικοποίηση υψηλότερου σχήματος διαμόρφωσης (Modulation Coding Scheme - MCS). Ο χειριστής του δικτύου επιδιώκει πάντα να μεγιστοποιεί το SINR σε όλες τις τοποθεσίες για να προσφέρει την καλύτερη δυνατή Εμπειρία Χρήστη, και αυτό επιτυγχάνεται είτε με μετάδοση σε υψηλότερη ισχύ είτε ελαχιστοποιώντας τις παρεμβολές και τον θόρυβο. Για τους σκοπούς αυτή της προσομοίωσης χρησιμοποιούμε την πιο κάτω εξίσωση για την εκτίμηση του SINR που μας δίνεται από το 3GPP [17]:

$$SINR_{i,j} = \frac{RSRP_{i,j}}{I_{i,j} + N_j} \quad (2)$$

Όπου I είναι η ισχύς της παρεμβολής των άλλων σημάτων στο δίκτυο που παρεμβάλουν, και N η ισχύς του θορύβου, που μπορεί να είναι μια σταθερά ή κάποια τυχαία μεταβλητή (π.χ. λευκός θόρυβος όταν ακολουθεί Gaussian κατανομή). Η ισχύς του λαμβανόμενου σήματος αναφοράς (Reference Signal Receive Power – RSRP) υπολογίζεται πάλι χρησιμοποιώντας το πρότυπο από το 3GPP ως:

$$RSRP_{i,j} = P_j + G_j - L_j - L_{i,j} \quad (3)$$

Όπου P_j είναι η ένταση ισχύος της αντένας, G_j το κέρδος της αντένας, L_j είναι οι απώλειες ενέργειας και $L_{i,j}$ είναι η απώλεια διαδρομής (Path Loss – PL), όπου χρησιμοποιείται το μοντέλο υπολογισμού βασισμένο στην απόσταση.[19]

$$L_{i,j} = 128.1 + 37.6 * \log_{10} distance \quad (4)$$

Υπολογισμος Throughput

Μια καλή τιμή SINR έχει κρίσιμο ρόλο στην μεγιστοποίηση της χωρητικότητας και της μετάδοσης δεδομένων (Throughput). Ακόμα κι αν δύο συνδρομητές μπορούν να χρησιμοποιούν την ίδια ποσότητα φάσματος (δηλαδή τον αριθμό OFDMA - Subcarriers), η αντίστοιχη ποιότητα σήματος (SINR) καθορίζει την απόδοσή τους. Αυτό οφείλεται στο γεγονός ότι όσο υψηλότερο είναι το SINR, τόσο μεγαλύτερη είναι η διαμόρφωση QAM που μπορεί να επιτευχθεί, και επομένως τόσο υψηλότερος είναι ο ρυθμός δεδομένων που θα βιώσει ο συνδρομητής. Στις ζώνες χαμηλότερης συχνότητας και σε πολυάσχολες τοποθεσίες κυψέλης, ένα χαμηλό SINR έχει υπερβολικό αρνητικό αντίκτυπο, επομένως η βελτιστοποίηση SINR είναι ιδιαίτερα σημαντική σε αυτές τις περιπτώσεις. Έτσι μπορούμε να θεωρήσουμε ότι η κλιμάκωση του SINR έχει άμεσο αντίκτυπο στην διαμόρφωση που θα χρησιμοποιηθεί. Ως εκ τούτου επιλέγουμε για τους σκοπούς της προσομοίωσης να αντιστοιχίσουμε σε κάθε UE_i συνδεδεμένο σε ένα Σταθμό Βάσης BS_j μια διαμόρφωση σύμφωνα με τον πάρακατω πίνακα 2 που προσφέρεται από το σύμφωνο 3GPP:

Το MCS ορίζει τον αριθμό των χρήσιμων bit που μπορούν να μεταφερθούν από ένα σύμβολο. Η Σειρά Μορφοποίησης που θα επιλεγεί λαμβάνοντας υπόψη το SINR και τον δείκτη Κατάστασης καναλιού (Channel Quality Indicator– CQI) προκύπτει από τον **Πίνακα 1** [15], και θα χρησιμοποιηθεί από το Φυσικό Κανάλι Κατερχόμενης Ζεύξης (Physical Downlink Shared Channel - PDSCH). Η μετάδοση στο δίκτυο 5G εξαρτάται από ένα στοιχείο, ένα σύμβολο που ορίζεται ως Στοιχείο πόρου (Resource Element - RE) και το MCS ορίζει ως πόσα χρήσιμα bit μπορούν να μεταδοθούν ανά RE. Το MCS εξαρτάται από την ποιότητα του ραδιοφωνικού σήματος στην ασύρματη σύνδεση, όσο καλύτερη η ποιότητα τόσο υψηλότερο είναι το MCS και τόσο περισσότερα τα χρήσιμα bit που μπορούν να μεταδοθούν σε ένα σύμβολο, η κακή ποιότητα σήματος οδηγεί σε χαμηλότερο MCS σημαίνει ότι μπορούν να μεταδοθούν λιγότερα χρήσιμα δεδομένα σε ένα σύμβολο. Έτσι ορίζοντας το SINR έχουμε μια ένδειξη της ποιότητας της σύνδεσης μεταξύ του UE και του BS.

Λαμβάνοντας υπόψη τη μετάδοση κατερχόμενης ζεύξης (Downlink) σε ένα δίκτυο 5G πολλών χρηστών, όπου το μοντέλο OFDM χρησιμοποιείται για τη μετάδοση δεδομένων μέσω πολλών υποφορέων (Sub-carriers) στενής ζώνης (30 kHz) σε αριθμό n χρηστών. Στον τομέα συχνότητας, κάθε 12 SCs ομαδοποιούνται για να σχηματίσουν το βασικό στοιχείο προγραμματισμού του 3GPP 5G που ονομάζεται μπλοκ πόρων (Resource Block).

CQI	MCS	Code rate × 1024	Spectral efficiency	SNR (dB)	
				Perfect channel estimation	Practical channel estimation
1	QPSK	78	0.1523	-11.2	-6.3
2	QPSK	120	0.2344	-6.9	-5.8
3	QPSK	193	0.377	-2.2	-1.4
4	16QAM	308	0.6016	2.7	3.9
5	16QAM	449	0.877	4.3	5.3
6	16QAM	602	1.1758	6.9	8.1
7	64QAM	378	1.4766	8.5	9.8
8	64QAM	490	1.9141	10.6	11.7
9	64QAM	616	2.4063	12.4	13.6
10	64QAM	466	2.7305	14.4	15.8
11	64QAM	567	3.3223	17.5	18.8
12	256QAM	666	3.9023	18.1	21.4
13	256QAM	772	4.5234	20.2	23.6
14	256QAM	873	5.1152	22.8	28.2
15	256QAM	948	5.5547	24.9	32

Πίνακας 1. CQI and SINR mapping

Table 5.1.3.1-2: MCS index table 2 for PDSCH

MCS Index I_{MCS}	Modulation Order Q_m	Target code Rate $R \times [1024]$	Spectral efficiency
0	2	120	0.2344
1	2	193	0.3770
2	2	308	0.6016
3	2	449	0.8770
4	2	602	1.1758
5	4	378	1.4766
6	4	434	1.6953
7	4	490	1.9141
8	4	553	2.1602
9	4	616	2.4063
10	4	658	2.5703
11	6	466	2.7305
12	6	517	3.0293
13	6	567	3.3223
14	6	616	3.6094
15	6	666	3.9023
16	6	719	4.2129
17	6	772	4.5234
18	6	822	4.8164
19	6	873	5.1152
20	8	682.5	5.3320
21	8	711	5.5547
22	8	754	5.8906
23	8	797	6.2266
24	8	841	6.5703
25	8	885	6.9141
26	8	916.5	7.1602
27	8	948	7.4063
28	2	reserved	
29	4	reserved	
30	6	reserved	
31	8	reserved	

Πίνακας 2. MCS index Table

Έτσι στην συνέχεια χρησιμοποιώντας την παρακάτω εξίσωση, που ακολουθείται από το σύμφωνο 3GPP 38.306 [18] παίρνουμε μια χονδρική εκτίμηση του ρυθμού μετάδοσης δεδομένων για ένα δοσμένο αριθμό aggregated carriers:

$$\text{Data Rate (Mbps)} = 10^{-6} \sum_{j=1}^J v_{Layers}^{(j)} * Q_m^{(j)} * f^{(j)} * R_{max} * \frac{N_{PRB}^{BW(j),\mu} * 12}{T_s^\mu} * (1 - OH^{(j)}) \quad (5)$$

J : αριθμός Φορέων Συστατικών (Component Carriers) σε συνδυασμό με την χρησιμοποιημένη ζώνη ή συνδυασμό ζωνών.

$v_{Layers}^{(j)}$: Αριθμός MIMO layers. Σύμφωνα με το 3GPP 38.802 μέγιστος αριθμός για DL είναι 8, μέγιστος αριθμός για UL είναι 4.

$Q_m^{(j)}$: Σειρά Μορφοποίησης. Ανάλογα την μορφοποίηση έχουμε την αντίστοιχη μετατροπή συμβόλων σε bits. Σύμφωνα με τον Πίνακα 2 παίρνει τιμές 2/4/6/8 και αντιστοιχεί σε QPSK/16QAM/64QAM/256QAM.

$f^{(j)}$: Παράγοντας Κλιμάκωσης. Προκύπτει από τον συνδυασμό των CCs επιλεγούμε να το ορίσουμε σύμφωνα με τον πίνακα από 3GPP TS 38.306.

Table 1. Scaling factors for some band combination configurations

		SCS (kHz)	Maximum CBW (MHz)	Maximum number of Layers	Maximum modulation order	Real Maximum data rate	Calculated with maximum values <u>from</u> <u>BPC</u> and band combination	Scaling factor
<i>BPC</i>		<i>N/A</i>	<i>100</i>	<i>4</i>	<i>8</i>	<i>N/A</i>	<i>N/A</i>	<i>N/A</i>
CA_1	CC1	30	100	4	8	R1	R1	1
CA_2	CC1	30	100	2	8	R1	2*R1	0.5
	CC2	30	100	2	8			
CA_3	CC1	30	100	4	6	3/4*2*R1	2*R1	0.75
	CC3	30	100	4	6			

Πίνακας 3. 5G NR Scaling Factor

R_{max} : Η τιμή εξαρτάται από τον τύπο κωδικοποίησης του 3GPP 38.212 που φαίνεται στον πίνακα της Εικόνας 16.

$N_{PRB}^{BW(j),\mu}$: Ο μέγιστος αριθμός RBs για τον συγκεκριμένο συνδυασμό Bandwidth, Subcarrier Spacing και 5G Numerology $\mu(i)$ όπως προκύπτει από τον πιο κάτω πίνακα που δίνεται από το 3GPP 38.211.

Slot length scales with the subcarrier spacing: Slot length = 1 ms/2 ^{μ}

Numerology	Subcarrier spacing	# slots per subframe	Slot length
0	15 kHz	1	1ms/2 ⁰ =1 ms
1	30 kHz	2	1ms/2 ¹ =500us
2	60 kHz	4	1ms/2 ² =250 us
3	120 kHz	8	1ms/2 ³ =125μs

Πίνακας 4. 5G NR Numerology

T_s^μ : Η μέση διάρκεια του συμβόλου OFDM στο υποπλαίσιο για την τιμή $\mu(i)$ για το κανονικό κυκλικό πρόθεμα (Cyclic Prefix), η οποία υπολογίζεται σύμφωνα με την Εξίσωση:

$$T_s^\mu = \frac{10^{-3}}{14 * 2^\mu} \quad (6)$$

$OH^{(j)}$: Το Overhead που λαμβάνει κάποια από τις παρακάτω τιμές ανάλογα την ζώνη και το είδος της ζεύξης.

0.14 Για FR1 και DL

0.18 Για FR2 και DL

0.08 Για FR1 και UL

0.10 Για FR2 και UL

Για να λάβουμε ένα ακριβές αποτέλεσμα, είναι απαραίτητο να εισαγάγουμε τις σημαντικές παραμέτρους δικτύου (λειτουργία δικτύου 5G, αριθμός CCs, αριθμός επιπέδων MIMO, ζώνη συχνοτήτων, τύπος διαμόρφωσης). Οι παράμετροι του προσομοιωμένου δικτύου φαίνονται στον παρακάτω πίνακα [22].

Parameter	Setting
Macrocells	Number varies
Air Protocol	5G NR
5G Frequency Range	FR 1
RB Bandwidth	360 KHz
Modulation Scheme	2/4/6/8 (Varies based on reported SINR)
Bandwidth	100 MHz
Max Number of RBs	273
Subcarrier Spacing	30 KHz
Number of SCs Per RB	12
Cycle Prefix	Normal
Macrocell Coverage	500m
BS Antenna Gain	15 dBi
UE Antenna Gain	0 dBi
MIMO Layers	4
OFDM Symbols	14
Slot Duration	0.5 ms
DL:UL Ratio	4:1

Πίνακας 5. Παράμετροι Δικτύου

Έτσι μπορούμε να συμπεράνουμε πως το τελικό Throughput που θα έχει κάθε χρήστης ανα χρονική στιγμή εξαρτάται από τους μεταβαλλόμενους παράγοντες που είναι το $Q_m^{(j)}$, R_{max} , και ο αριθμός των RBs που εξαρτάται από τις απαιτήσεις του από το δίκτυο.

Υπολογισμός Απόστασης – Μετρική Haversine

Όπως έχουμε ήδη αναφέρει για τους σκοπούς αυτής της διπλωματικής έχουμε να κάνουμε με πραγματικές γεωγραφικές συντεταγμένες αποτυπωμένες στην μορφή Γεωγραφικού πλάτους (Latitude) και Γεωγραφικού μήκους (Longitude). Αυτό μας εμποδίζει από το να χρησιμοποιήσουμε μετρικές απόστασης που θα χρησιμοποιούσαμε αν είχαμε να κάνουμε με τον ευκλείδειο χώρο. Για αυτό τον λόγο για την μέτρηση της απόστασης μεταξύ δύο σημείων χρησιμοποιούμε την φόρμουλα Haversine, η οποία μας επιτρέπει να υπολογίσουμε την γεωγραφική απόσταση στη Γη. Η φόρμουλα Haversine πρακτικά μας επιτρέπει να υπολογίσουμε την μικρότερη απόσταση μεταξύ δύο σημείων στην επιφάνεια μιας σφαίρας με τριγωνομετρική προσέγγιση.

$$a = \sin^2\left(\frac{Latitude_{point1} - Latitude_{point2}}{2}\right) + \cos(Latitude_{point1}) * \cos(Latitude_{point2}) * \sin^2\left(\frac{Longitude_{point1} - Longitude_{point2}}{2}\right)$$

$$c = 2 * \text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$Distance = Earth\ Radius * c$$

3.2 Συσταδοποίηση Χρηστών

Στην συνέχεια θα παρουσιάσουμε τον μηχανισμό συσταδοποίησης χρηστών για τον προγραμματισμό κελιών το οποίο μπορεί να χρησιμοποιηθεί σε χονδρική εκτίμηση του αριθμού των κυττάρων. Επιπλέον, τα δεδομένα που λαμβάνονται μπορεί να είναι χρήσιμα σε αρχικά επενδυτικά σχέδια όπου η γνώση του αριθμού κυψέλης είναι ζωτικής σημασίας για τον τελικό επιχειρηματικό αντίκτυπο. Ο σωστός και λεπτομερής σχεδιασμός του ραδιοφωνικού δικτύου είναι ένα σημαντικό στάδιο στο σχεδιασμό ολόκληρου του δικτύου, επειδή το ραδιοδίκτυο μπορεί να έχει μεγάλο αντίκτυπο στο κόστος υλοποίησης, στη βελτιστοποίηση των θέσεων των σταθμών βάσης, στη χρησιμοποιούμενη ισχύ, στα ύψη της κεραίας και σε άλλες πρακτικές πτυχές. Η απόδοση της επένδυσης μπορεί να διαφέρει σημαντικά. Ακόμη η σωστή τοποθέτηση των χρηστών είναι μια απαραίτητη ενέργεια για την εύρεση της βέλτιστης τοποθεσίας των Σταθμών Βάσης αλλά και των αριθμό που χρειαστούν για να καλύψουν τις ανάγκες των χρηστών σε ρυθμό δεδομένων. Στόχος μας είναι η εύρεση της τοποθεσίας των BS ώστε να μεγιστοποιήσουμε τον αριθμό των χρηστών που εξυπηρετούνται και ως εκ τούτου να ελαχιστοποιήσουμε τον αριθμό των χρηστών που δεν εξυπηρετούνται επειδή δεν βρίσκονται στην εμβέλεια κάποιου BS.[19][20]

```
def calculateCenters(thisData):
    R = 6378.1 # Radius of the Earth
    d = 0.2 # Distance in km
    global centers2
    global NUMBER_OF_CLUSTERS
    global bsTable
    global algo_name
    global time_clusters

    lat = []
    lon = []
    dict = {}

    for i in range(0, len(thisData)):
        lat.append(float(thisData[i]['lat']))
        lon.append(float(thisData[i]['lon']))
    first = np.column_stack((lat, lon))
    # scaler = RobustScaler()
    # first = scaler.fit_transform(first)

    #ELBOW METHOD
    # K_clusters = range(1,10)
    # kmeans = [KMeans(n_clusters=i) for i in K_clusters]
    # score = [kmeans[i].fit(first).inertia_ for i in range(len(kmeans))]
    # # Visualize
    # # print(score)
    # # plt.plot(K_clusters, score)
    # # plt.xlabel('Number of Clusters')
    # # plt.ylabel('Score')
    # # plt.title(f'Elbow Curve, {algo_name}, Number of users = {len(first)}')
    # # plt.show()

    t = time.time()
    if algo_name == 'pam':
        kmedoids = KMedoids(n_clusters=NUMBER_OF_CLUSTERS, random_state=0, metric='haversine', method='pam').fit(first)
        labels = kmedoids.predict(first)
        centers = kmedoids.cluster_centers_
    elif algo_name == 'kmeans':
        kmean = KMeans(n_clusters=NUMBER_OF_CLUSTERS, init='k-means++')
        labels = kmean.fit_predict(first)
        centers = kmean.cluster_centers_
    elif algo_name == 'birch':
        birch = Birch(threshold=0.0058, n_clusters=NUMBER_OF_CLUSTERS)
        labels = birch.fit_predict(first)
        centers = birch.subcluster_centers_
    elif algo_name == 'dbscan':
        dbscan = DBSCAN(eps=0.00175, min_samples=5, algorithm='auto', metric='haversine').fit(np.radians(first))
        labels = dbscan.fit_predict(first)
        new = np.column_stack((first, labels))
        df = pd.DataFrame(new, columns=['lat', 'lon', 'label'])
        centers_lat = df.groupby(['label'])['lat'].mean()
        centers_lon = df.groupby(['label'])['lon'].mean()
        centers = []
        for x,y in zip(centers_lat, centers_lon):
```

Κώδικας 6. Κώδικας Δημιουργίας BS (part 1)

```

        centers.append([x,y])
        centers = np.array(centers)
    elif algo_name == 'myalgo':
        df = pd.DataFrame(first, columns=['lat', 'lon'])
        centers_lat = df['lat'].mean()
        centers_lon = df['lon'].mean()
        centers = [[centers_lat, centers_lon]]
        centers = np.array(centers)

    time_clusters = round(time.time() - t, 4)
    print(f"Clustering was performed in: {time_clusters} sec")
    # centers = scaler.inverse_transform(centers)
    centers2 = centers.copy()

    # clustering_analysis(first, labels, centers)

    for center in centers2:
        bsTable.append(BaseStation(center[0], center[1]))
    centers = centers.tolist()
    print(centers2)

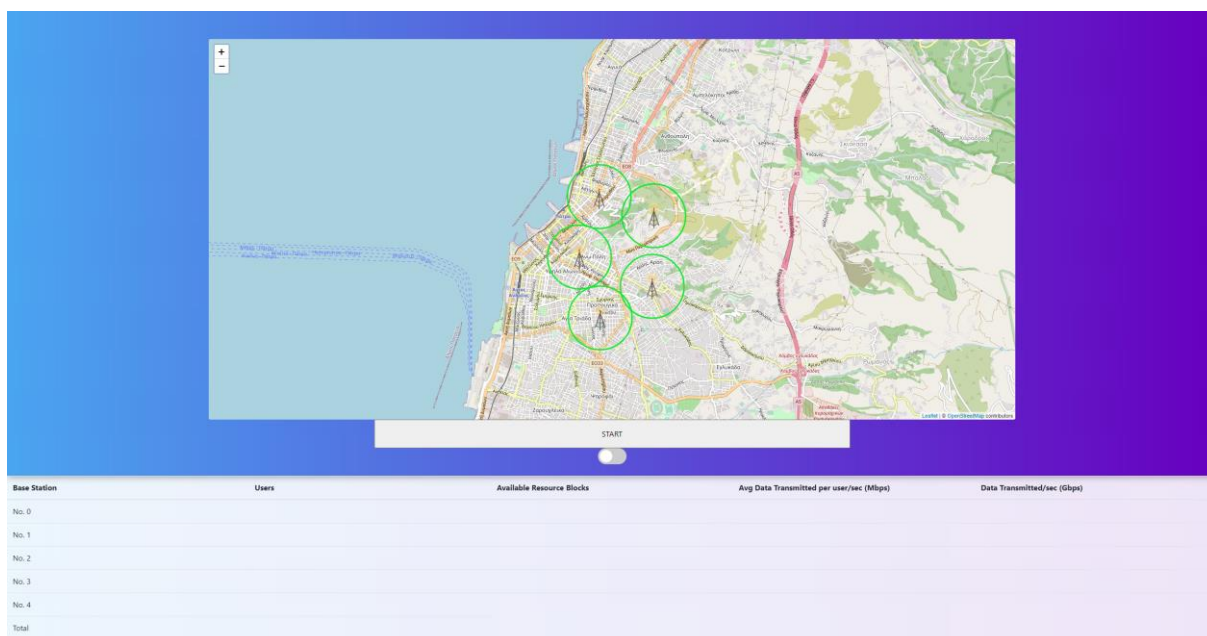
    for i in range(0, len(centers)):
        dict[i] = [centers[i], newPoint(R, 0, d, centers[i]), newPoint(R, 1.57, d, centers[i]), newPoint(R, 3.14, d, centers[i]), newPoint(R, 4.71, d, centers[i])]
    return dict

```

Κώδικας 7. Κώδικας Δημιουργίας BS (part2)

Η δημιουργία των κέντρων αποτελεί το πρώτο κομμάτι κώδικα που τρέχει στην μεριά του server και τρέχει μόνο την πρώτη φορά για τον υπολογισμό της βέλτιστης τοποθεσίας των BSs, με τέτοιο τρόπο που θα καλύψουμε όσο το δυνατό περισσότερους χρήστες. Το παραπάνω κομμάτι κώδικα τρέχει προτού ξεκινήσει η προσομοίωση και το έναυσμα δίνεται με ένα ajax request από την μεριά του frontend όπου στέλνονται οι γεωγραφικές συντεταγμένες των τυχαίων σημείων που αντιπροσωπεύουν τα UE και τροφοδοτούνται στον εκάστοτε αλγόριθμο συσταδοποίησης. Μέσω του κώδικα είναι δυνατή η δημιουργία διαφόρων σεναρίων όσο αφορά το είδος της συσταδοποίησης (K-means, PAM, Birch) το πλήθος των BSs, και το πλήθος των UEs. Μέσα από πειράματα έχει διαπιστωθεί πως η προσομοίωση είναι ικανή να υποστηρίξει μέχρι και περιβάλλοντα με 10.000 UEs όπου η θέση τους μεταβάλλεται συνεχώς μαζί με τα επιμέρους στοιχεία τους. Όσο αυξάνουμε τον αριθμό των UEs, αυξάνεται ραγδαία και το μέγεθος του http request, στο σημείο η καθυστέρηση που προκαλείται να μειώνει την εμπειρία της προσομοίωσης. Γι' αυτό και τα αποτελέσματα που θα παρουσιάσουμε το επόμενο κεφάλαιο δεν ξεπερνάνε τα 1600 UEs.

Πρώτη ενέργεια είναι η μετατροπή των δεδομένων σε μορφή κατάλληλη για να τροφοδοτηθούν στον αλγόριθμο συσταδοποίησης. Στην συνέχεια έχουμε την επιλογή να χρησιμοποιήσουμε κάποιον scaler για να κανονικοποιήσουμε τα δεδομένα. Ωστόσο καθώς τα δεδομένα μας βρίσκονται όλα στην ίδια κλίμακα με μικρή διασπορά βλέπουμε ελάχιστη βελτίωση επί τους αποτελέσματος και έτσι επιλέγουμε να μην το χρησιμοποιήσουμε για να αποφύγουμε το επιπλέον overhead. Μέσα σε ένα if block κάθε φορά ελέγχουμε ποιον αλγόριθμο ομαδοποίησης έχουμε θέσει ως όρισμα και τροφοδοτούμε σε αυτόν τα δεδομένα. Από κάθε αλγόριθμο παίρνουμε τα παραγόμενα κέντρα. Δημιουργούμε τα στιγμιότυπα της κλάσης BaseStation που παρουσιάσαμε στο προηγούμενο κεφάλαιο, δίνοντας ως όρισμα τις γεωγραφικές συντεταγμένες των κέντρων κάθε BS και επιστρέφουμε ως απάντηση στο ajax request τα κέντρα των BSs που θα χρησιμοποιηθούν από την εφαρμογή για απεικόνιση στον χάρτη.



Εικόνα 19. Start Screen of the Simulation

3.3 Συσχετισμός Χρηστών με Σταθμούς Βάσης

Σύμφωνα με τις τεχνικές προδιαγραφές για τα 5G Δίκτυα από το πρωτόκολλο διεπαφής ραδιοφώνου NR (Third Generation Partnership Project, 2018a), στο φυσικό επίπεδο του δικτύου, η συμβατική Ορθογώνια Πολυπλεξία Διαίρεσης Συχνότητας (Orthogonal Frequency Division Multiplexing - OFDM) χρησιμοποιείται για το δίκτυο DL με κανονικό κυκλικό πρόθεμα. Θεωρούμε ότι όλοι οι Σταθμοί Βάσης (στο συγκεκριμένο πείραμα χειριζόμαστε MacroCells) μπορούν να ικανοποιήσουν έναν χρήστη, εφόσον θεωρείται ως το βέλτιστο BS για να τον εξυπηρετήσει. Όλα τα καθορισμένα Base Stations έχουν περιορισμένους πόρους και κατά συνέπεια μπορούν να ικανοποιήσουν μόνο περιορισμένο αριθμό χρηστών ταυτόχρονα. Όλα τα BS του ίδιου τύπου έχουν τους ίδιους διαθέσιμους πόρους, αλλά ο αριθμός των χρηστών που μπορούν να εξυπηρετήσουν ποικίλλει ανάλογα με τις μετρήσεις μεμονωμένων χρηστών. Παράγοντες που επηρεάζουν είναι RSRP, RSRQ, Path Loss και το SINR (Signal to Noise Ratio) όπου και θα χρησιμοποιήσουμε στο παράδειγμα. Ανάλογα με το SINR προκύπτει και ο αριθμός CQI που θα μας καθορίσει το Modulation Coding Scheme που θα χρησιμοποιηθεί, άρα έτσι μπορούμε να βρούμε το μέγιστο Throughput που μπορεί να προκύψει για ένα χρήστη.[21]

Όταν ένα UE_j επιλέγει να συσχετιστεί με ένα BS_i , για αυτό το UE υπάρχει η πιθανότητα να διαλέξει ένα BS που έχει τον μικρότερο αριθμό RBs σε σχέση με τα υπόλοιπα του δικτύου ή και που βρίσκονται στην εμβέλεια του, αλλά μπορεί να ικανοποιήσει τις απαιτήσεις για την δημιουργία σύνδεσης. Οι απαιτήσεις σε RBs του UE είναι ανάλογες με τις απαιτήσεις διακίνησης και αντιστρόφως ανάλογες με το εύρος ζώνης ενός συγκεκριμένου RB και του SINR μεταξύ του UE και του BS. Η εξίσωση για τον αριθμό των RBs που απαιτούνται για μια συσχέτιση UE-BS προέρχεται από το θεώρημα του Shannon [7] και υπολογίζεται ως:

$$RB_{S_{j,i}} = \left\lceil \frac{T_j}{B_{RB} * \log_2(1 + SINR_{j,i})} \right\rceil$$

όπου $\lceil \cdot \rceil$ is είναι ο τελεστής για τη λειτουργία οροφής, T_j δηλώνει τις απαιτήσεις του UE σε Throughput για την δραστηριότητα που εκτελεί, B_{RB} αντιστοιχεί στο εύρος ζώνης ενός συγκεκριμένου RB και $SINR_{j,i}$ την ποιότητα ασύρματης σύνδεσης μεταξύ του UE και του BS.

Η απόδοση κάθε χρήστη συνδέεται άρρηκτα με τον αριθμό των RB που απαιτεί και το SINR του BS στο οποίο συνδέεται, επομένως όσο υψηλότερο είναι το SINR, τόσο υψηλότεροι θα είναι οι ρυθμοί δεδομένων και τόσο καλύτερη κωδικοποίηση των δεδομένων μπορεί να προκύψει. Δεδομένου ότι η πλειονότητα των UE επιλέγει να συνδεθεί σε ένα κελί MacroCell, κάθε χρήστης πιθανότατα θα συνδεθεί στο BS στο οποίο έχει το καλύτερο SINR και έτσι θα αποκτήσει υψηλούς ρυθμούς δεδομένων. Όμως, όσο περισσότερους χρήστες προσθέτουμε στο δίκτυο, τόσο δυσκολότερο γίνεται για τους χρήστες να βρουν ένα βέλτιστο BS για συσχέτιση καθώς αυξάνεται η ισχύς της παρεμβολής

από τα υπόλοιπα UE του BS και ως εκ τούτου μειώνεται το SINR προς τον επιθυμητό BS. Ο θόρυβος πρόκειται για μια τυχαία μεταβλητή που ακολουθεί την Κανονική Κατανομή. Αυτό μπορεί να οδηγήσει ένα UE στο να συνδεθεί σε ένα BS που βρίσκεται σε μεγαλύτερη απόσταση από τη βέλτιστη, εάν η αρχική προσπάθεια να δημιουργήσει μια βέλτιστη συσχέτιση μεταξύ UE και BS δεν αποτύχει. Τελικά, βλέπουμε ότι οι μέσοι ρυθμοί δεδομένων χρήστη τείνουν να μειώνονται εάν προσθέσουμε περισσότερους χρήστες στο δίκτυο και το πυκνώσουμε. Ωστόσο, καθώς το UC model (User Centric model) υπόσχεται επιπλέον περισσότερη ελευθερία λόγω της δυνατότητας σύνδεσης σε διαφορετικά BS κατά την διάρκεια του Downlink και του Uplink, γνωστό και ως Downlink and Uplink Decoupling (DUD), για καλύτερη ποιότητα σύνδεσης, βρισκόμαστε αντιμέτωποι με υψηλότερους ρυθμούς δεδομένων που θα είχαμε εάν το δίκτυο ακολουθούσε μια προσέγγιση Network Centric (NC). Όσο για τους συνολικούς ρυθμούς δεδομένων δικτύου, παρουσιάζουν εκπληκτική αύξηση, καθώς υπολογίζονται ως το άθροισμα όλων των Data Rates για κάθε UE τόσο στα δίκτυα κατερχόμενης όσο και άνω ζεύξης.

Για κάθε χρήστη πρώτα κοιτάζουμε σε ποιους Σταθμούς Βάσης βρίσκεται εντός εμβέλειας, και στην συνέχεια κατατάσσουμε τους Σταθμούς Βάσης σύμφωνα με το εκτιμώμενο throughput που μπορούν να προσφέρουν στον χρήστη στην περίπτωση που μπορεί να μονοπωλήσει όλους τους πόρους. Υπολογίζουμε τα απαιτούμενα RBs για την επίτευξη του επιθυμητού throughput βάση των στοιχείων του δικτύου όπως δείξαμε στην εξίσωση (5). Στην συνέχεια βάση της σειράς κατάταξης προσπελαύνουμε τους Σταθμούς βάσης και αναθέτουμε τον χρήστη σε αυτόν τον σταθμό που μπορεί να εξυπηρετήσει τις ανάγκες του σε RBs κοιτάζοντας τα δικά του διαθέσιμα RBs που δεν καταλαμβάνονται από άλλους χρήστες, και έπειτα υπολογίζουμε το throughput βάση των απαιτούμενων RBs.

Ο αλγόριθμος στοχεύει στο να μεγιστοποιήσουμε τον αριθμό των χρηστών που εξυπηρετούνται εντός της εμβέλειας του Macrocell, αλλά αυτό επιτυγχάνεται στο κόστος του να μείνουν μερικοί χρήστες με υψηλές απαιτήσεις μη εξυπηρετημένοι. Ο αλγόριθμος λαμβάνει υπόψη τις απαιτήσεις των χρηστών σε RBs και προσπελαύνει τα UE ταξινομημένα με αύξουσα σειρά ως προς τις απαιτήσεις τους με σκοπό να εξυπηρετήσει όσους περισσότερους χρήστες το δυνατόν. Απαραίτητα για την εκτέλεση του αλγορίθμου είναι η γνώση μετρικών όπως τα διαθέσιμα RBs του BS, τις απαιτήσεις κάθε χρήστη σε RBs και το SINR από κάθε BS ως προς κάθε UE για τον υπολογισμό των τιμών CQI και το είδος της μορφοποίησης που θα χρησιμοποιηθεί. Με επαναληπτικό τρόπο κάθε UE διαλέγει το BS που μπορεί αν του προσφέρει τον καλύτερο ρυθμό δεδομένων μέχρι να βρεθεί ένα τέτοιο BS, διαφορετικά το UE μένει μη εξυπηρετημένο.

Τον αλγόριθμο για τον συσχετισμό χρηστών με Σταθμούς Βάσης στην απλοϊκή του μορφή μπορούμε να τον δούμε παρακάτω.

Αλγόριθμος 1. Συσχέτιση Χρηστών με BS

N: Αριθμός UE

For i = 1 to N **do**

Υπολόγισε το πλήθος των Σταθμών Βάσης στον οποίον βρίσκεται στην εμβέλεια

If Πλήθος Σταθμών Βάσης > 2 **Then****For** j=1 to Πλήθος Σταθμών Βάσης **do****If** Διαθέσιμα RBs >= Απαιτούμενα RBs **Then**

Υπολόγισε το Throughput του UE ως προς τον Σταθμό Βάσης

End For

Ανάθεσε τον χρήστη στον Σταθμό Βάσης με το μεγαλύτερο Throughput

Else If Πλήθος Σταθμών Βάσης = 1 **do**

Ανάθεσε το UE στον Σταθμό Βάσης

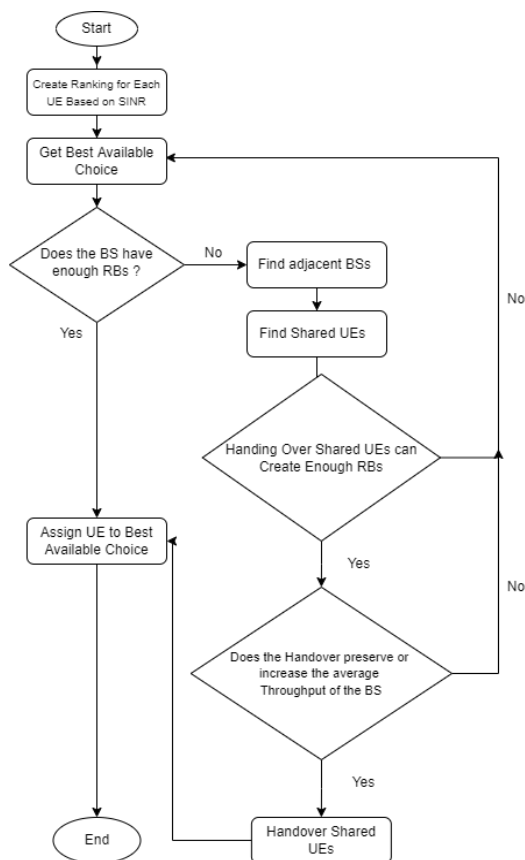
Else If Πλήθος Σταθμών Βάσης = 0 **do**

Το UE δεν βρίσκεται στην Εμβέλεια κάποιου Σταθμού Βάσης, μην κάνεις τίποτα

End For

Για να αντιμετωπίσουμε το πρόβλημα των χρηστών με μεγάλες απαιτήσεις, που ως επί των πλείστων έχαναν την σειρά τους από τα UEs που μονοπωλούσαν τους πόρους του δικτύου λόγω των μικρών τους απαιτήσεων, αναπτύχθηκε ένας αλγόριθμος τύπου Handover. Ο αλγόριθμος ξεκινάει παρόμοια με την απλοϊκή του μορφή, ωστόσο κάθε φορά που κάποιος BS τον απορρίπτει επειδή δεν έχει τους απαραίτητους πόρους για να τον εξυπηρετήσει τότε αντί να στέλνει ένα ACK αίτημα στον επόμενο διαθέσιμο BS (αν υπάρχει) τότε ο BS προσπαθεί να παραχωρήσει χρήστες σε διπλανούς BSs για να απελευθερώσει αρκετούς πόρους ώστε να εξυπηρετήσει το νέο UE. Πρώτα ο BS αναγνωρίζει τα UE που βρίσκονται στην εμβέλεια άλλων BSs και τα ταξινομεί με φθίνουσα σειρά. Στην συνέχεια υπολογίζει ποιο συνδυασμό αυτών των UEs θα χρειαστεί να παραδώσει στους γειτονικούς BSs για να απελευθερώσει αρκετούς πόρους. Ο BS στέλνει ένα καταναμημένο μήνυμα προς τους Σταθμούς Βάσης που τέμνονται σε αυτόν για να του απαντήσουν με τους διαθέσιμους τους πόρους. Αν οι πόροι είναι αρκετοί τότε εκτελείται το Handover των UEs διαφορετικά το UE μένει μη εξυπηρετημένο, και συνεχίζει με τον αμέσως καλύτερο BS σύμφωνα με το εκτιμώμενο SINR.

Ο προαναφερόμενος μηχανισμός συσχέτισης χρηστών με Σταθμούς Βάσης περιγράφεται στο παρακάτω ακολουθιακό διάγραμμα, μαζί με τον μηχανισμό Handover δικής μας επινόησης που στοχεύει στην μεγιστοποίηση του αριθμού χρηστών που εξυπηρετούνται.



Σχήμα 1. Μηχανισμός Handover

Φυσικά αυτό είναι ένα πρόβλημα που σε πραγματικές συνθήκες δεν θα εμφανιζόταν συχνά καθώς η παράταξη των μοντέρνων δικτύων 5ης Γενιάς είναι σε μορφή HetNet όπου μικρότερες κυψέλες υπάρχουν εντός της εμβέλειας, (συνήθως στα όρια της Μακροκυψέλης όπου οι χρήστες αντιμετωπίζουν χειρότερη σύνδεση επειδή το σήμα εξασθενεί ή επειδή δέχονται παρεμβόλες από γειτονικές Κυψέλες) που είναι σε θέση να προσφέρουν καλύτερο Throughput και σε περίπτωση που δεν μπορούν τότε η εξωτερική κυψέλη θα καλύψει τις ανάγκες του UE.

4

ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

4.1 Σύνοψη

Σε αυτό το Κεφάλαιο θα παρουσιάσουμε τις τεχνικές που χρησιμοποιήθηκαν για την επιλογή του αριθμού των Συστάδων (οι όροι Σταθμός Βάσης και Συστάδα χρησιμοποιούνται αλληλένδετα), για την εκτίμηση της ποιότητας των Συστάδων και επιλογή του καλύτερου αλγόριθμου για τον υπολογισμό των κέντρων. Στην συνέχεια θα παρουσιάσουμε μερικά στατιστικά αποτελέσματα πάνω στον αριθμό των UE που καταφέρνουν οι αλγόριθμοι να καλύψουν σε διάφορα στιγμιότυπα του πειραματικού περιβάλλοντος που περιγράψαμε στο προηγούμενο Κεφάλαιο (3). Τέλος θα δούμε πως κυμαίνεται ο μέσος ρυθμός μετάδοσης δεδομένων βάση των Σταθμών Βάσης που παρήγαγε ο καλύτερος αλγόριθμος συσταδοποίησης και πως ο μηχανισμός Handover που περιγράψαμε αποδίδει σε σχέση με την απλοϊκή λύση.

Είναι σημαντικό να έχουμε κατά νου ότι υπάρχει ένα αυστηρό όριο στο πόσα δεδομένα μπορούν να μεταδοθούν σε ένα δεδομένο εύρος ζώνης σύμφωνα με το θεώρημα Shannon-Hartley (Shannon's Law):

$$C = n * B * \log_2(1 + SINR)$$

C = Χωρητικότητα του Καναλιού (bits/sec)

N = Αριθμός από αντένες

B = Bandwidth (Hz)

Αυτό το θεώρημα ορίζει τον μέγιστο ρυθμό με τον οποίο μπορούν να μεταδοθούν πληροφορίες μέσω ενός καναλιού επικοινωνίας, μέσω του αέρα στην περίπτωση μας, ενός καθορισμένου εύρους ζώνης παρουσία θορύβου. Καθορίζει ένα όριο στη μέγιστη ποσότητα πληροφοριών χωρίς σφάλματα ανά μονάδα χρόνου που μπορεί να μεταδοθεί με ένα συγκεκριμένο εύρος ζώνης παρουσία της παρεμβολής θορύβου, υποθέτοντας ότι η ισχύς του σήματος είναι περιορισμένη και ότι η διαδικασία του Gaussian θορύβου χαρακτηρίζεται από ένα γνωστό φασματική πυκνότητα ισχύος.

4.2 Συσταδοποίηση Χρηστών

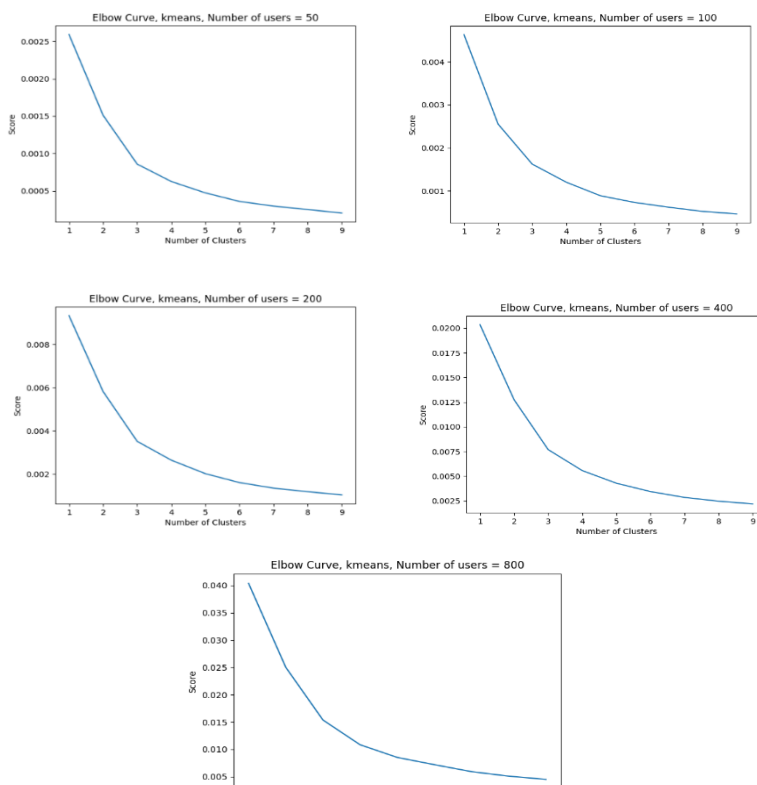
Στην συνέχεια θα παρουσιάσουμε μερικές τεχνικές ως προς την εκτίμηση της ποιότητας των συστάδων κάθε αλγορίθμου που χρησιμοποιήθηκε, τον λόγο που επιλέχθηκε ο αριθμός των συστάδων και τα αποτελέσματα επί των αριθμών των χρηστών που καταφέραμε να καλύψουμε αλλά και το υπολογιστικό κόστος κάθε αλγορίθμου.

4.2.1 Elbow Method

Οι αλγόριθμοι συσταδοποίησης που χρησιμοποιήθηκαν στο πειραματικό κομμάτι K-means και K-medoids(PAM) ως αλγόριθμοι τμηματοποίησης χρειάζεται να τους ορίσουμε τον αριθμό των συστάδων βάση του οποίου θα τμηματοποιήσουν τον χώρο για να δημιουργήσουν τις συστάδες. Από την άλλη ο αλγόριθμος Birch ως ιεραρχικός αλγόριθμος χρειάζεται να ρυθμίσουμε ανάλογα την τιμή Threshold βάση της οποίας θα προκύψουν οι συστάδες. Συχνά το πρόβλημα εύρεσης του αριθμού των συστάδων δεν έχει ντετερμινιστικό τρόπο λύσης, και εξαρτάται από τον τύπο του προβλήματος που έχουμε να λύσουμε. Στην δική μας περίπτωση οι συστάδες αναπαριστούν Σταθμούς Βάσης, και για αυτό τον λόγο δεν έχουμε την ευχέρεια του να τοποθετήσουμε αυθαίρετα μεγάλο αριθμό συστάδων καθώς η τοποθέτηση ενός BS κοστίζει τόσο σε χρήματα αλλά και ενέργεια. Επιπλέον όσο αυξάνουμε τον αριθμό των BS τόσο μεγαλύτερη γίνεται και η ισχύς της παρεμβολής μεταξύ των BS που έχει ως αποτέλεσμα να παίρνουμε χειρότερες τιμές SINR άρα και να οδηγούμαστε σε χειρότερους ρυθμούς μετάδοσης δεδομένων. Για αυτό τον λόγο θέλουμε να βρούμε τον βέλτιστο αριθμό συστάδων, όπου από αυτό τον αριθμό και πάνω δεν βλέπουμε μεγάλη βελτίωση στον αριθμό χρηστών που καταφέρνουμε να εξυπηρετήσουμε αλλά ταυτόχρονα καταφέρνουμε να καλύψουμε ένα ικανοποιητικό ποσοστό χρηστών επί του συνόλου.

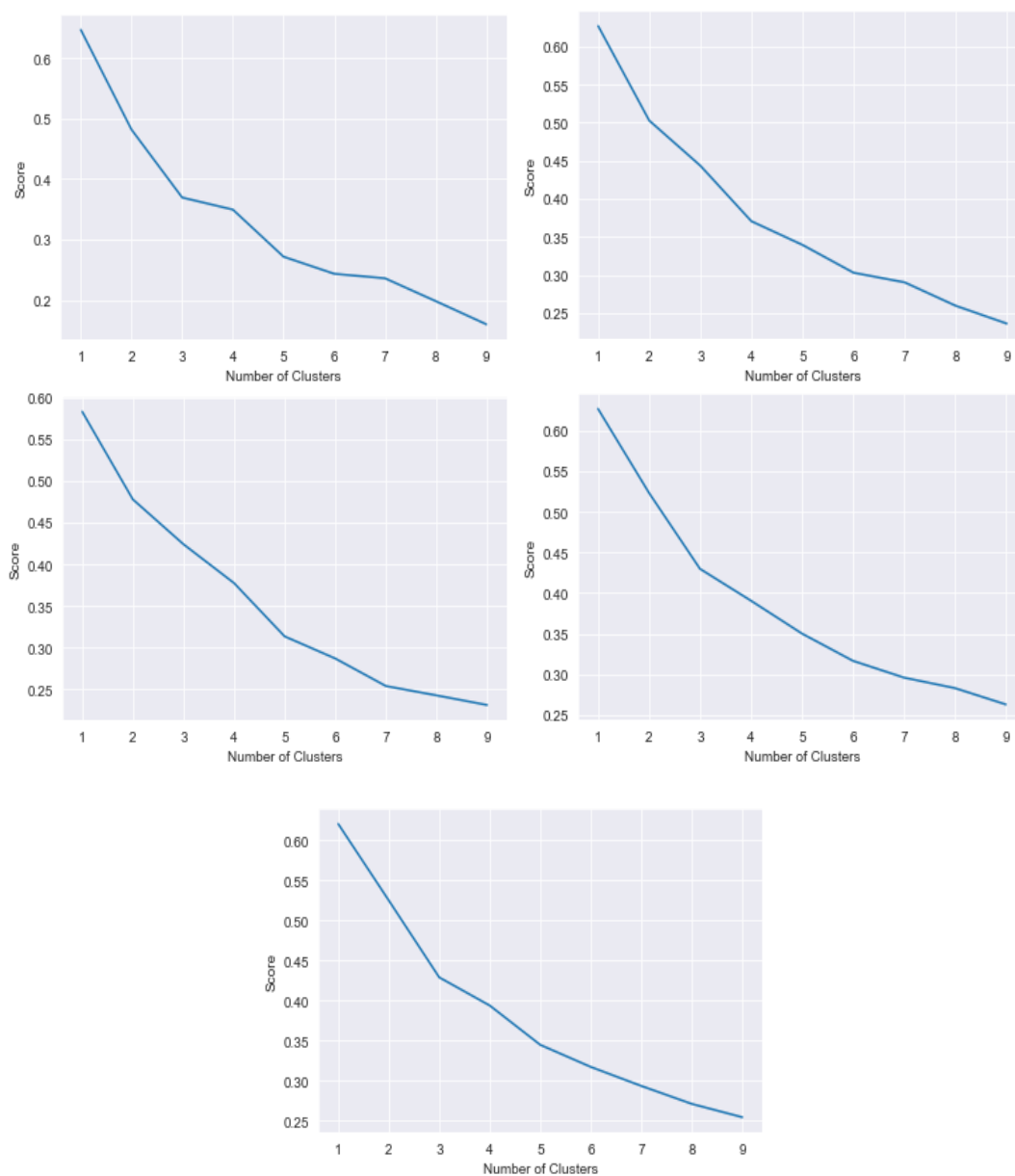
Για να καταλήξουμε στον βέλτιστο αριθμό συστάδων χρησιμοποιούμε την Elbow Method. Αυτή η μέθοδος χρησιμοποιεί το άθροισμα της τετραγωνικής απόστασης (SSE) για να επιλέξει μια ιδανική τιμή k με βάση την απόσταση μεταξύ των σημείων και των κέντρων της συστάδας που τους έχει εκχωρηθεί. Σαν κανόνα επιλέγουμε μια τιμή k όπου το SSE αρχίζει να δείχνει ένα σημείο καμπής. Όταν οπτικοποιηθεί αυτό το γράφημα θα μοιάζει κάπως με αγκώνα, εξού και το όνομα της μεθόδου.[37]

K – means



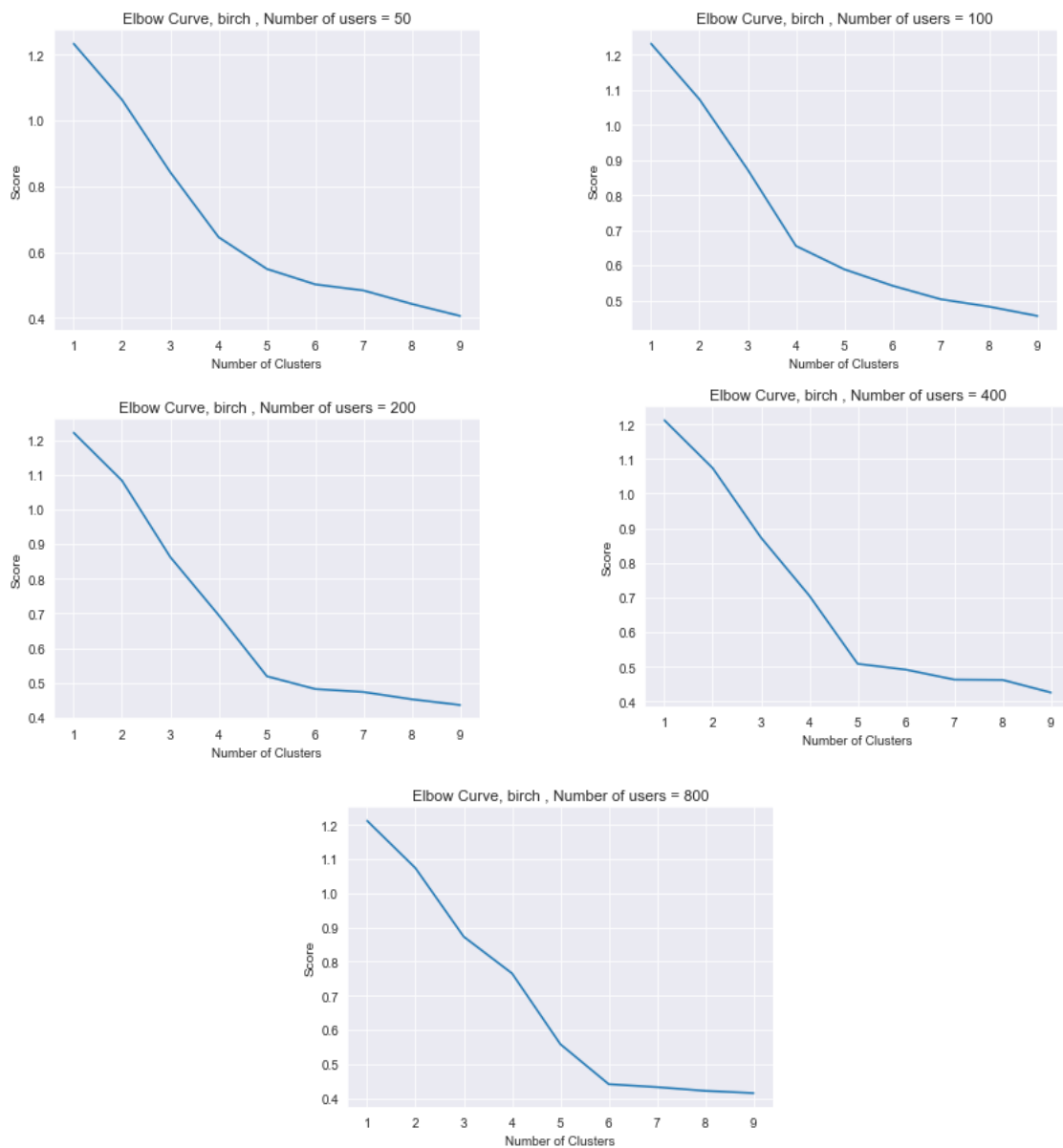
Σχήμα 2. Elbow Method K-means

K – Medoids (PAM)



Σχήμα 3. Elbow Method K-Medoids(PAM)

Birch



Σχήμα 4. Elbow Methods Birch

Από τα γραφήματα μπορούμε να δούμε ότι στην πλειονότητα των περιπτώσεων η μέση απόσταση μετά από 4-5 συστάδες ξεκινάει να μειώνεται χωρίς αξιοσημείωτη βελτίωση. Από τη στιγμή που κάθε συστάδα αντιπροσωπεύει ένα Σταθμό Βάσης, η απόσταση των σημείων από το κέντρο που τους ανατέθηκε δεν μειώνεται αρκετά για να μας οδηγήσει στο να χρησιμοποιήσουμε πέρα των 5 συστάδων.

4.2.2 Silhouette Score

Η βαθμολογία Silhouette χρησιμοποιείται για την αξιολόγηση της ποιότητας των συστάδων που δημιουργήθηκαν χρησιμοποιώντας αλγόριθμους ομαδοποίησης όπως ο K-Means, K-medoids(PAM) και Birch. Στην πράξη αυτή η μετρική αφορά το πόσο καλά τα δείγματα ομαδοποιούνται με άλλα δείγματα που είναι παρόμοια μεταξύ τους. Η βαθμολογία Silhouette υπολογίζεται για κάθε δείγμα διαφορετικών συστάδων με τον εξής τρόπο:

$$\text{Silhouetter Score} = (b - a) / \max(a,b)$$

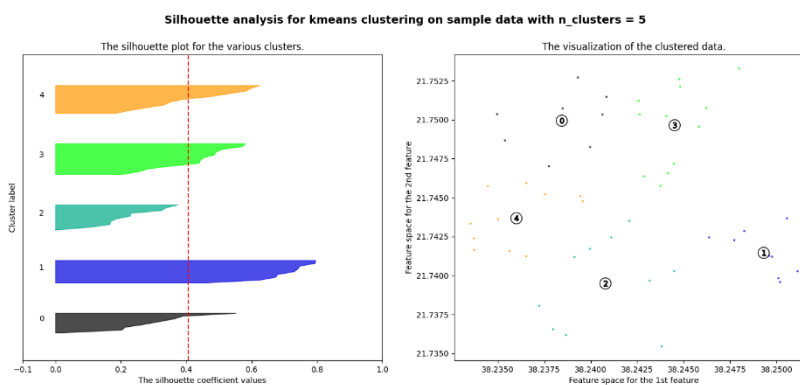
Όπου:

a = Η μέση απόσταση μεταξύ του σημείου και όλων των υπόλοιπων σημείων στην συστάδα την οποία ανήκει

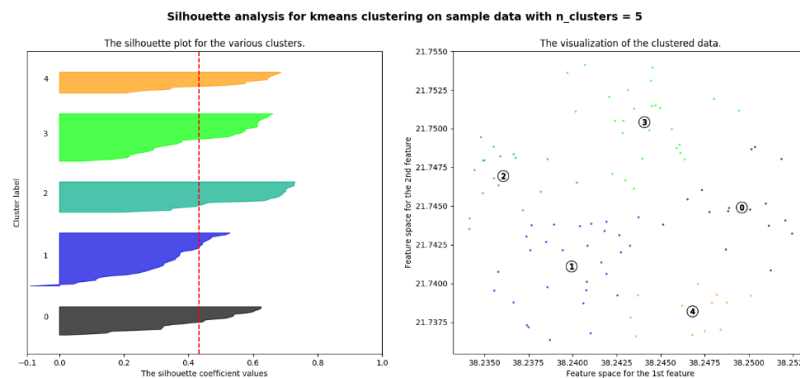
b = Η μέση απόσταση μεταξύ του σημείου και των υπόλοιπων σημείων από την κοντινότερη συστάδα στην οποία δεν ανήκει

Η τιμή της βαθμολογίας Silhouette κυμαίνεται από -1 έως 1. Εάν η βαθμολογία είναι 1, τότε η συστάδα είναι πυκνή και καλά διαχωρισμένη από τις υπόλοιπες. Μια τιμή κοντά στο 0 αντιπροσωπεύει επικαλυπτόμενες συστάδες με δείγματα πολύ κοντά στο όριο απόφασης των γειτονικών συστάδων. Μια αρνητική βαθμολογία υποδηλώνει ότι τα δείγματα μπορεί να έχουν εκχωρηθεί σε λάθος ομάδες.[36]

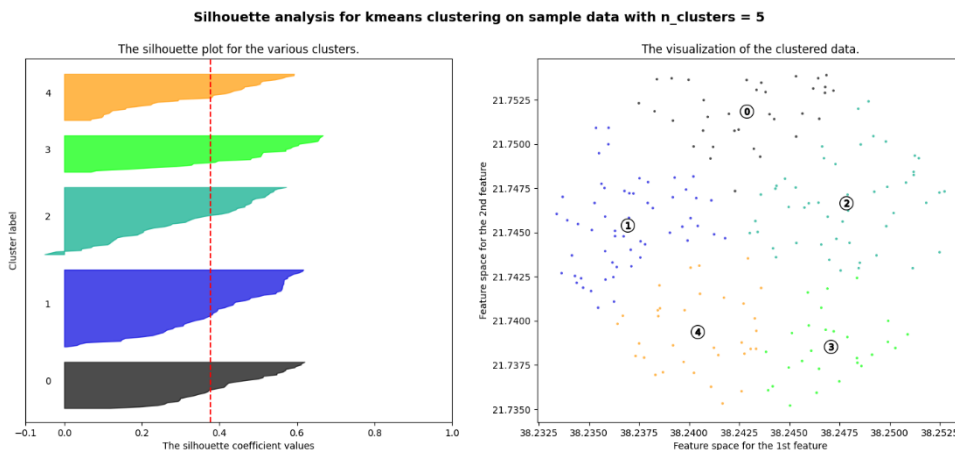
K – means



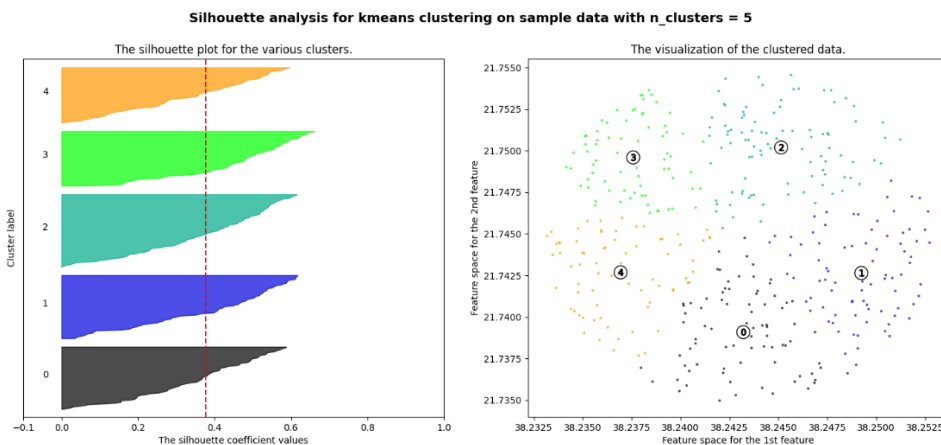
Σχήμα 5. K-means, users=50, Average silhouette_score is : 0.44138870374298655



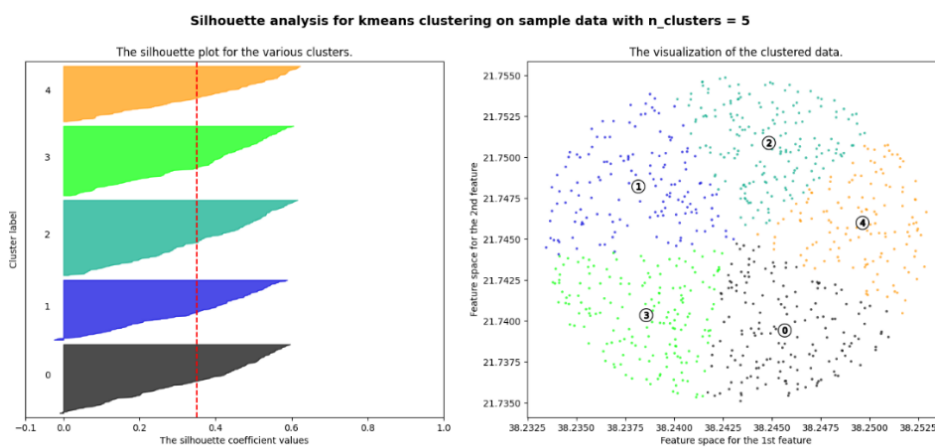
Σχήμα 6. K-means, users=100, Average silhouette_score is : 0.43380580197503243



Σχήμα 7. K-means, users=200, Average silhouette_score is : 0.3771023809015711

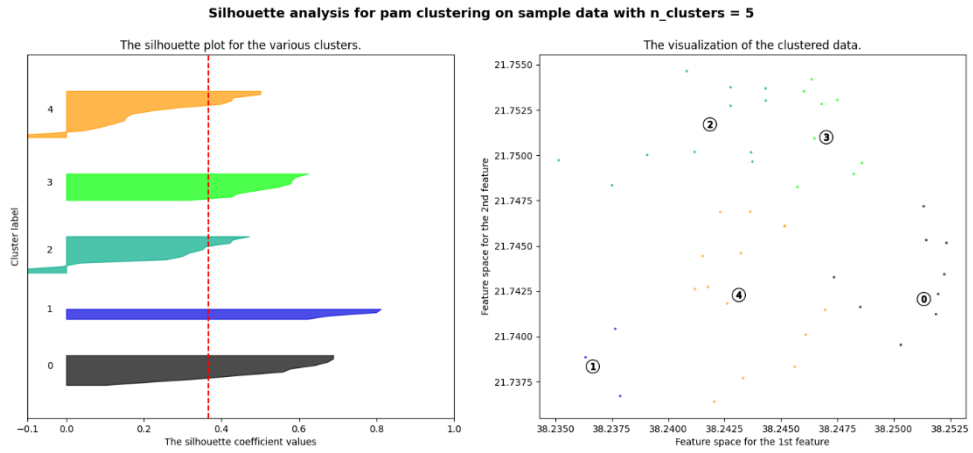


Σχήμα 8. K-means, users=400, Average silhouette_score is : 0.37622084556785224

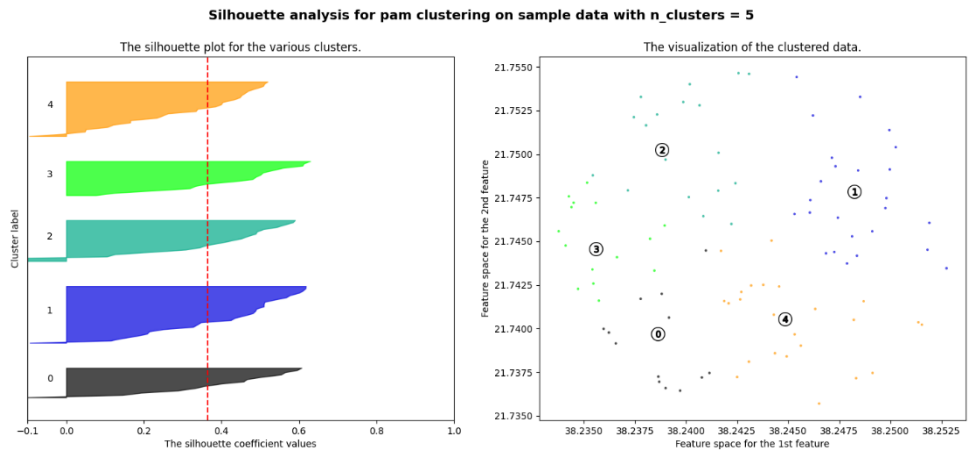


Σχήμα 9. K-means, users=800, Average silhouette_score is : 0.35068334190816214

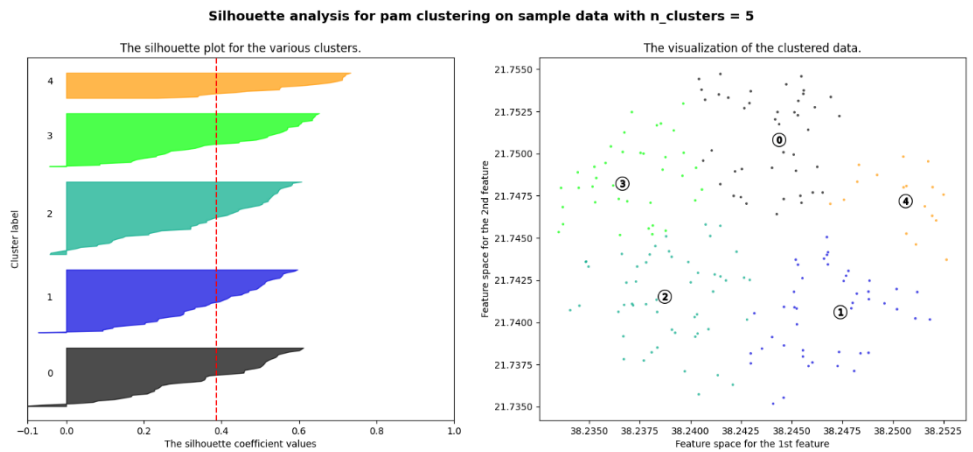
K – medoids (PAM)



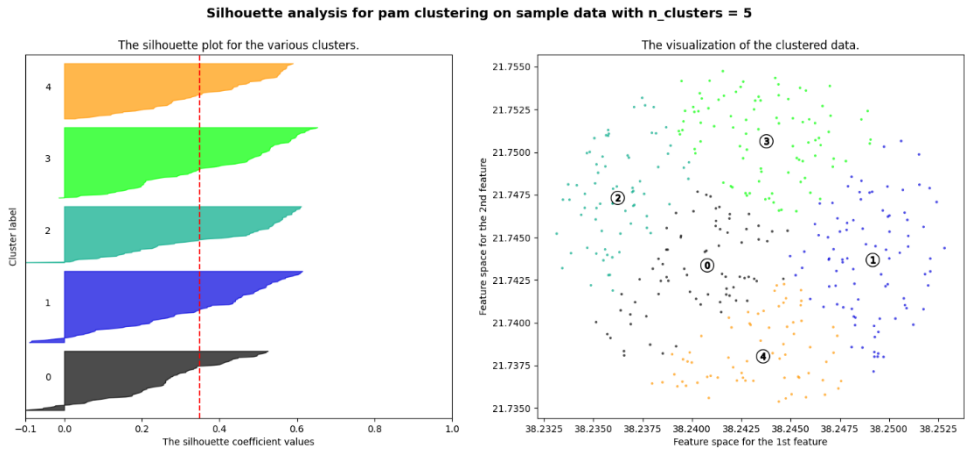
Σχήμα 10. PAM, users=50, The average silhouette_score is : 0.36619911250743536



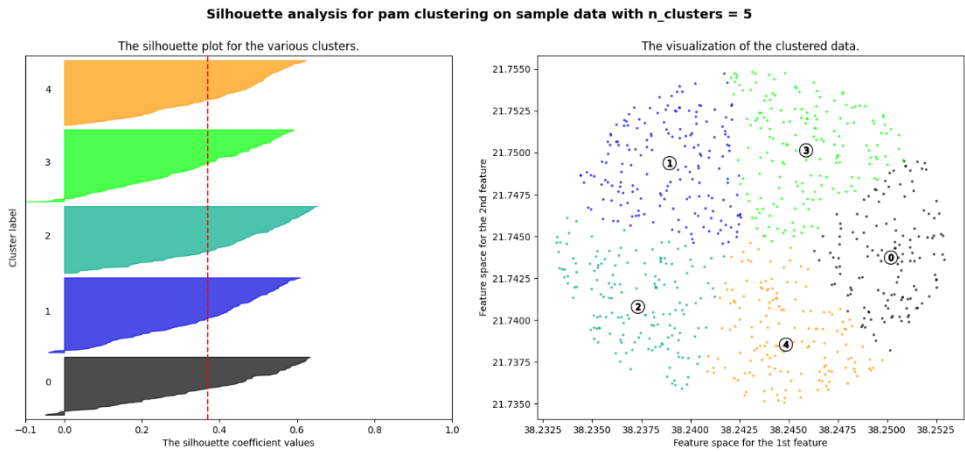
Σχήμα 11. PAM, users=100, Average silhouette_score is : 0.3652894167234125



Σχήμα 12. PAM, users=200, Average silhouette_score is : 0.3872290003478699

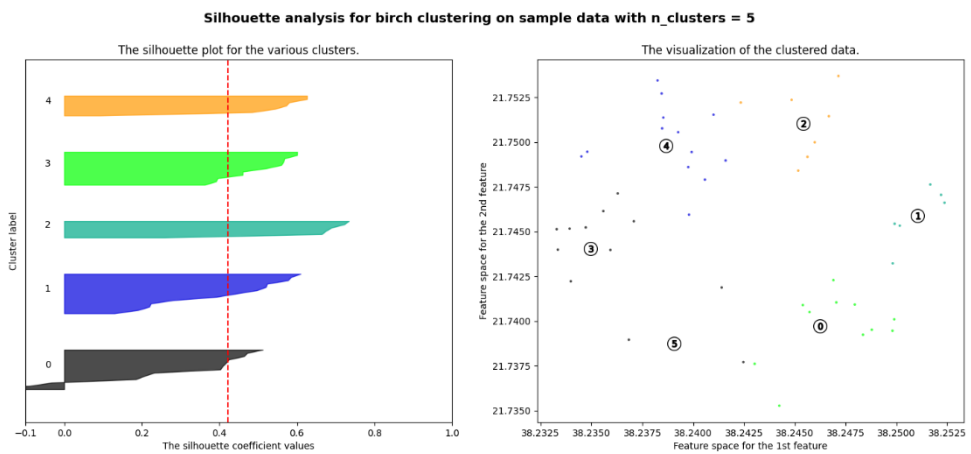


Σχήμα 13. PAM, users=400, Average silhouette_score is : 0.3494148478152864

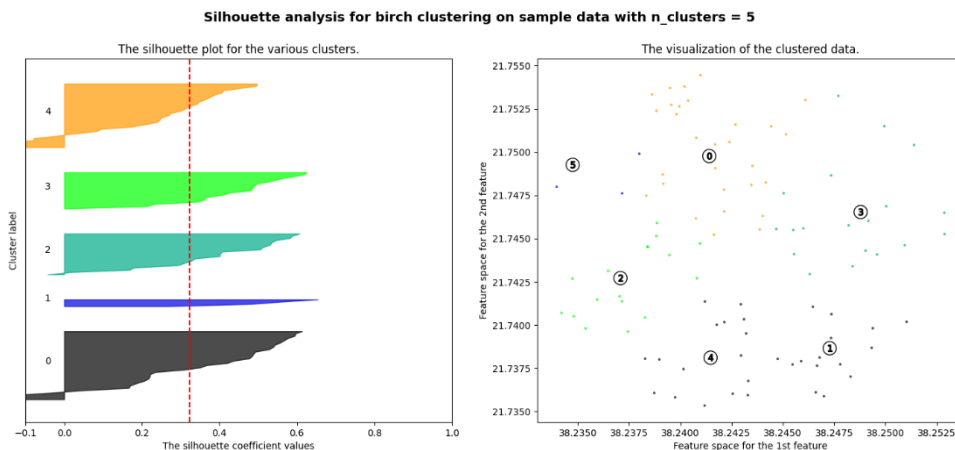


Σχήμα 14. PAM, users=800, Average silhouette_score is : 0.37031077201504997

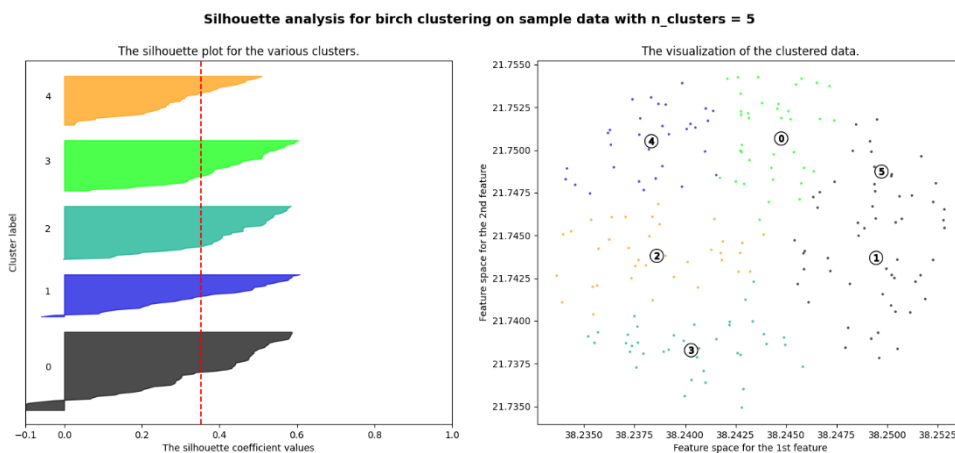
Birch



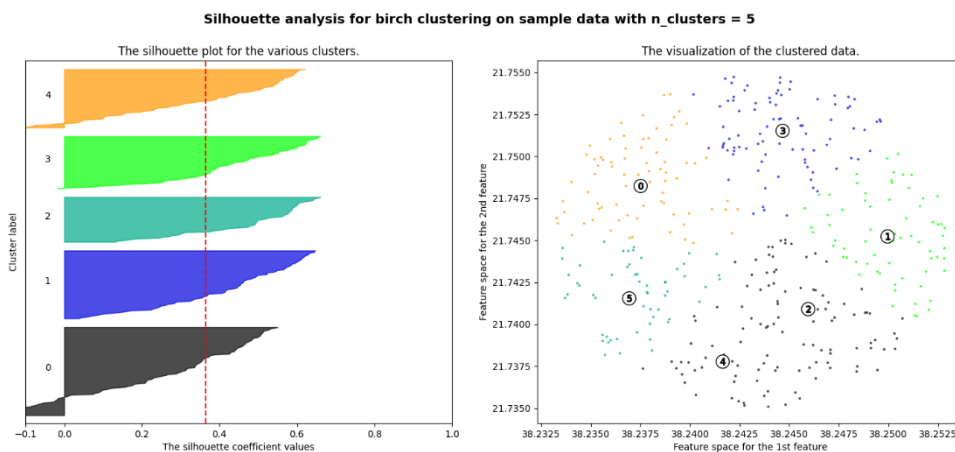
Σχήμα 15. Birch, users=50, Average silhouette_score is : 0.422720559467441



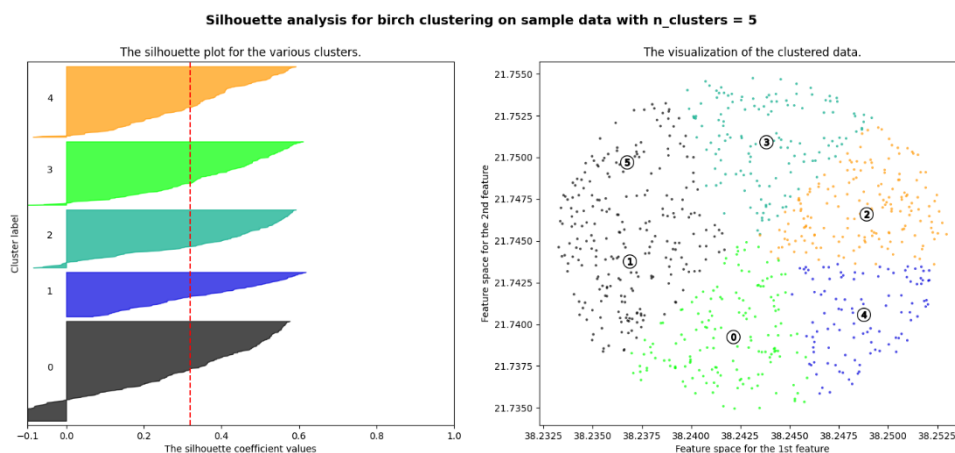
Σχήμα 16. Birch, users=100, Average silhouette_score is : 0.32339607482356636



Σχήμα 17. Birch, users=200, Average silhouette_score is : 0.3519887759110538



Σχήμα 18. Birch, users=400, Average silhouette_score is : 0.36518402227632424



Σχήμα 19. Birch, users=800, Average silhouette_score is : 0.3206565291438612

Όπως φαίνεται τα πειράματα έχουν γίνει για ένα σύνολο δεδομένων από 50 μέχρι 800 χρήστες με τον αριθμό να διπλασιάζεται κάθε φορά. Ο αριθμός των συστάδων είναι ο ίδιος για κάθε πείραμα, ορισμένος στις 5 συστάδες, όπως έδειξαν τα πειράματα με την elbow method που φαίνεται να είναι η βέλτιστη τιμή για την ομαδοποίηση.

Για αριθμό χρηστών $n = 50$ βλέπουμε ότι το καλύτερο average Silhouette Score προκύπτει από τον αλγόριθμο Kmeans, που έχει το υψηλότερο score και πιο κοντά στο 1. Δεύτερο καλύτερο Score είναι με τον αλγόριθμο Birch και τρίτο με τον αλγόριθμο PAM.

Για αριθμό χρηστών $n = 100$ βλέπουμε ότι το καλύτερο average Silhouette Score προκύπτει από τον αλγόριθμο Kmeans, που έχει το υψηλότερο score και πιο κοντά στο 1. Δεύτερο καλύτερο Score είναι με τον αλγόριθμο PAM και τρίτο με τον αλγόριθμο Birch.

Για αριθμό χρηστών $n = 200$ βλέπουμε ότι το καλύτερο average Silhouette Score προκύπτει από τον αλγόριθμο PAM, που έχει το υψηλότερο score και πιο κοντά στο 1. Δεύτερο καλύτερο Score είναι με τον αλγόριθμο Kmeans και τρίτο με τον αλγόριθμο Birch.

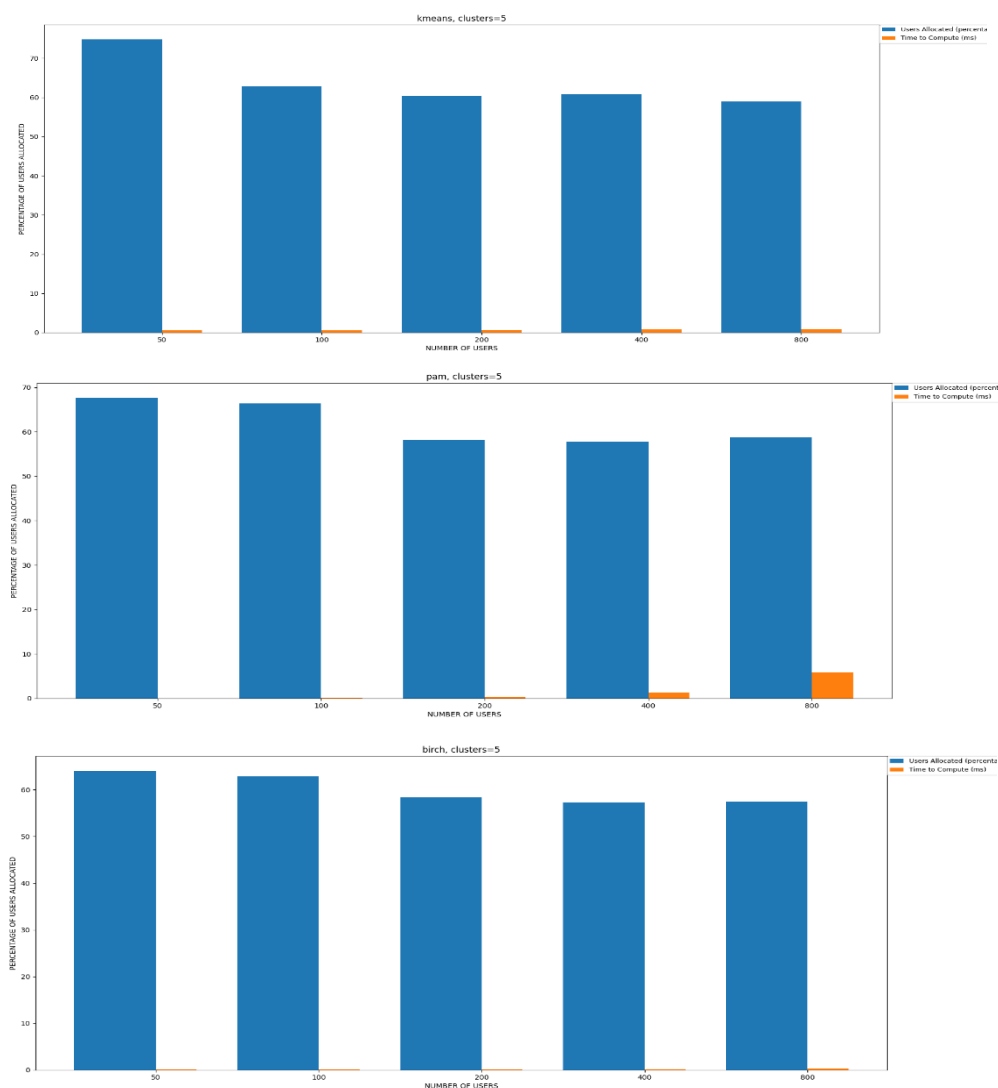
Για αριθμό χρηστών $n = 400$ βλέπουμε ότι το καλύτερο average Silhouette Score προκύπτει από τον αλγόριθμο Kmeans, που έχει το υψηλότερο score και πιο κοντά στο 1. Δεύτερο καλύτερο Score είναι με τον αλγόριθμο Birch και τρίτο με τον αλγόριθμο PAM.

Για αριθμό χρηστών $n = 800$ βλέπουμε ότι το καλύτερο average Silhouette Score προκύπτει από τον αλγόριθμο PAM, που έχει το υψηλότερο score και πιο κοντά στο 1. Δεύτερο καλύτερο Score είναι με τον αλγόριθμο Kmeans και τρίτο με τον αλγόριθμο Birch.

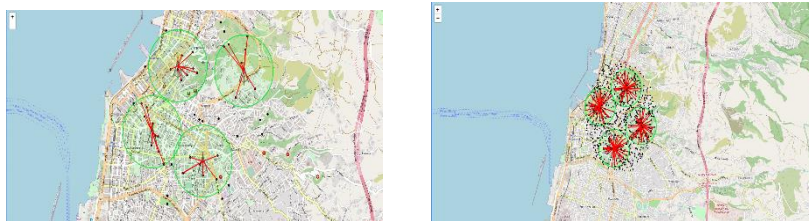
Γενικά μπορούμε να δούμε ότι σε μικρότερες τιμές χρηστών ο αλγόριθμος Kmeans προσφέρει καλύτερη διαχώριση μεταξύ των συστάδων ενώ όσο αυξάνεται ο αριθμός των χρηστών ο PAM φαίνεται να αποφέρει καλύτερα αποτελέσματα. Αυτό ίσως να συμβαίνει λόγω του ότι ο αλγόριθμος PAM υπολογίζει την απόσταση μεταξύ των σημείων χρησιμοποιώντας την μετρική Haversine ενώ ο Kmeans υπολογίζει την απόσταση μεταξύ των σημείων θεωρώντας ότι βρίσκονται στον Ευκλείδειο χώρο. Έτσι καθώς αυξάνεται ο αριθμός των σημείων και το δείγμα γίνεται πιο πυκνό είναι λογικό το average Silhouette Score να τείνει πιο κοντά στο 0 και ο αλγόριθμος PAM να έχει καλύτερα αποτελέσματα.

4.2.3 Συσχετισμός UE με BS

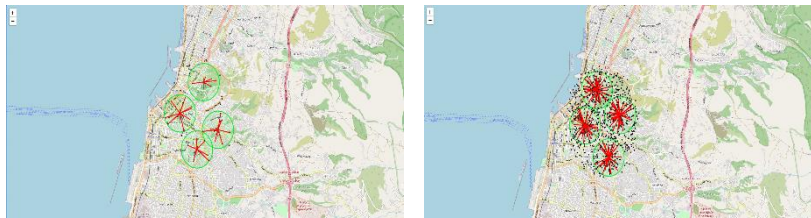
Στην συνέχεια παρουσιάζουμε το ποσοστό των χρηστών επι του συνόλου που κάθε αλγόριθμος κατάφερε να καλύψει και τον υπολογιστικό χρόνο που χρειάστηκε για την παραγωγή των κέντρων. Για κάθε αλγόριθμο τρέξαμε το πείραμα 100 φορές και πήραμε τον μέσο αριθμό από UE που βρίσκονται εντός της εμβέλειας κάθε Συστάδας/Σταθμού Βάσης (δηλαδή σε ακτίνα 500m για Macrocell) ανεξάρτητα αν αυτό το BS έχει αρκετούς πόρους για να εξυπηρετήσει όλα τα UE. Στόχος αυτής της ενότητας είναι να δείξουμε σε θέμα αποκλειστικά τοπολογικό πως ένας αλγόριθμος συσταδοποίησης είναι ικανός να παράγει πραγματικές γεωγραφικές τοποθεσίες Σταθμών Βάσης για την καλύτερη οργάνωση του Δικτύου. Προφανώς αυξάνοντας τον αριθμό των Σταθμών Βάσης ή την ακτίνα της εμβέλειας κάθε Σταθμού μπορούμε να φτάσουμε σε ένα σημείο 100% κάλυψης των χρηστών με το ανάλογο κόστος. Ωστόσο με αυτό το πείραμα μπορούμε να συμπεράνουμε σε συνθήκες περιορισμένων πόρων ποιος αλγόριθμος μπορεί να μας δώσει τα καλύτερα αποτελέσματα και κατ' επέκταση ποιον αλγόριθμο συμφέρει να χρησιμοποιήσουμε για να έχουμε καλύτερη κλιμάκωση του Δικτύου.



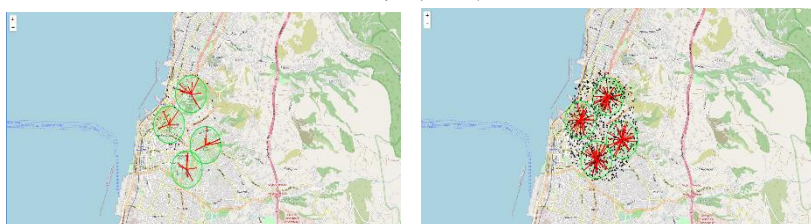
Σχήμα 20. Ποσοστά Κάλυψης UE



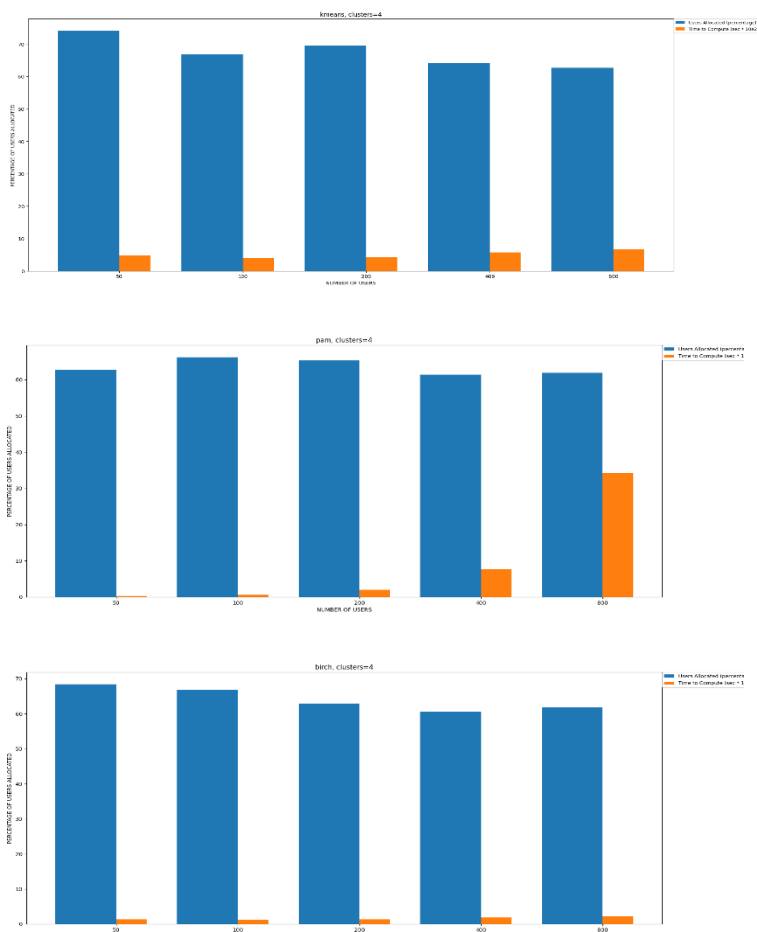
Εικόνα 20. Διάταξη BSs με K-means για $n=50$, $n=800$



Εικόνα 21. Διάταξη BSs με PAM για $n=50$, $n=800$



Εικόνα 22. Διάταξη BSs με Birch για $n=50$, $n=800$



Σχήμα 21. Ποσοστά Κάλυψης UE

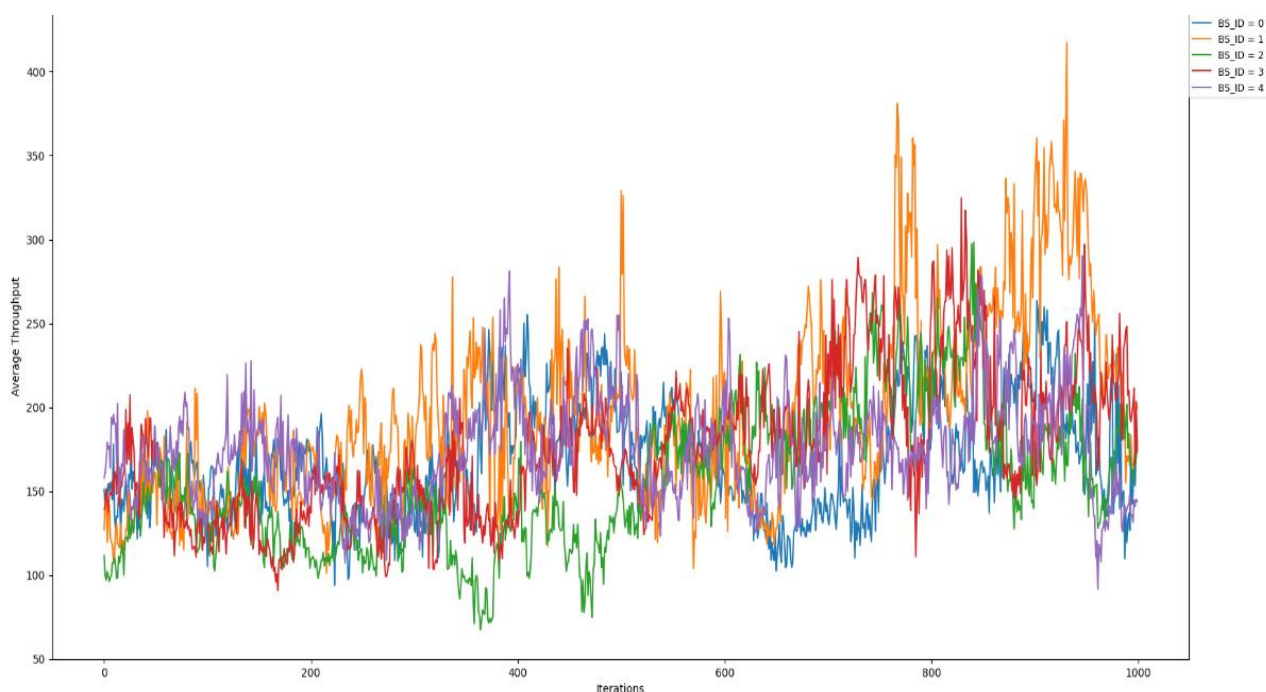
Απο τα πειράματα μπορούμε να δούμε ότι στην περίπτωση της χρήσης 4 Συστάδων/Σταθμών Βάσης για μικρό αριθμό χρηστών τα καλύτερα αποτελέσματα σε θέμα κάλυψης προκύπτουν από τον K-means με 73% κάλυψη, δεύτερος έρχεται ο ιεραρχικός αλγόριθμος Birch και τρίτος ο αλγόριθμος PAM. Καθώς αυξάνουμε τους χρήστες και φτάνουμε στους 400 το χάσμα μικραίνει, ωστόσο ο K-means εξακολουθεί να έχει την καλύτερη κάλυψη με 64%, δεύτερος αυτή την φορά ο PAM με 61% και τρίτος ο Birch με 60%. Στον μέγιστο αριθμό του πειράματος ο PAM καταφέρνει να ξεπεράσει τον K-means με ποσοστό 63% που συνεχίζει να βλέπει μια σταδιακή κάθοδο και φτάνει το 62% ενώ ο Birch παραμένει στο 60%. Οι τάσεις που μπορούμε να διακρίνουμε είναι ότι ο PAM από τους 50 χρήστες μέχρι τους 800 χρήστες βλέπει ελάχιστη διακύμανση καθώς σε κάθε σενάριο παραμένει μεταξύ 61% - 63% κάλυψη ενώ ο K-means βλέπει μια σταδιακή κάθοδο. Ο Birch συμπεριφέρεται παρόμοια με τον PAM, ωστόσο σε χαμηλότερα ποσοστά κάλυψης γύρω από το 60%. Αυτή η συμπεριφορά του PAM και του Birch μπορεί να οφείλεται στο ότι για τον υπολογισμό των κέντρων χρησιμοποιούν την μετρική Haversine, αντίθετα από τον K-means που χρησιμοποιεί την Ευκλείδεια απόσταση θεωρώντας ότι οι συντεταγμένες απεικονίζονται στον δισδιάστατο χώρο. Έτσι μπορούμε να συμπεράνουμε ότι στην περίπτωση δικτύου με μεγάλο αριθμό χρηστών (ένα πυκνό δίκτυο) η καλύτερη επιλογή είναι ο PAM. Ενώ για ένα αραιό δίκτυο η καλύτερη επιλογή είναι ο K-means.

Στην περίπτωση του χρόνου για 50 χρήστες ο γρηγορότερος αλγόριθμος είναι ο PAM με 3ms, δεύτερο ο Birch με 15 ms και τρίτος ο K-means με 50 ms. Ωστόσο καθώς ο αριθμός των σημείων αυξάνεται ο PAM αποδίδει χειρότερα λόγω της πολυωνυμικής του πολυπλοκότητας τάξεως $O(n^2)$, όπου στον μέγιστο αριθμό χρηστών φτάνει τα 350ms και αν συνεχίζαμε να διπλασιάζουμε τα σημεία, φτάνοντας στα 1600 πλέον θα χρειαζόταν δευτερόλεπτα. Ο Birch και ο K-means με την γραμμική χρονική πολυπλοκότητα τους φτάνουν χρόνους 25ms και 76ms στην περίπτωση των 800 σημείων, αρκετά καλή απόδοση που δεν γίνεται αντιληπτή.

Όσο για την τοπολογική διάταξη μπορούμε να δούμε πως όσο αυξάνεται ο αριθμός των σημείων και οι 3 αλγόριθμοι τείνουν να παράγουν χονδρικά τα ίδια κέντρα. Ωστόσο στην περίπτωση των 50 σημείων βλέπουμε για άλλη μια φορά πως ο PAM και ο Birch που χρησιμοποιούν την μετρική haversine συμπεριφέρονται παρόμοια, ενώ ο K-means ακολουθεί διαφορετική διάταξη.

4.2.4 Μελέτη ως προς την Βελτιστοποίηση του Ρυθμού Δεδομένων

Τώρα πλέον έχουμε υλοποιήσει το περιβάλλον προσομοίωσης και είμαστε σε θέση να δημιουργήσουμε τα BSs, UEs και να τα αντιστοιχίσουμε σε BSs χρησιμοποιώντας τον απλοϊκό αλγόριθμο Συσχέτισης Χρηστών με BS που δείξαμε στο Κεφάλαιο 3. Στα παρακάτω γραφήματα μπορούμε να δούμε τον μέσο ρυθμό μετάδοσης δεδομένων για κάθε BS που προκύπτει από την εξίσωση υπολογισμού Throughput που δείξαμε, και τον μέσο αριθμό UE που εξυπηρετούνται από κάθε BS, αυτή την φορά λαμβάνοντας υπόψη τον διαθέσιμο αριθμό από RBs για κάθε BS και τις απαιτήσεις κάθε UE. Οι τιμές έχουν δειγματοληπτηθεί ανα 100 επαναλήψεις, δηλαδή κάθε 100 αλλαγές τοποθεσιών κάθε χρήστη και επανυπολογισμού των επι μέρους στοιχείων που συνοδεύουν την αλλαγή κατάστασης ενός UE. Προτεραιότητα δίνεται στους χρήστες με τις μικρότερες απαιτήσεις σε RBs καθώς σκοπός είναι να μεγιστοποιήσουμε τον αριθμό των UE που εξυπηρετούνται. Για τις γραφικές έχουν χρησιμοποιηθεί 400 UE με 5 BSs, και σύμφωνα με το προηγούμενο υπο-Κεφάλαιο επιλέγουμε να χρησιμοποιήσουμε τον K-means που παρουσίασε τα καλύτερα αποτελέσματα.

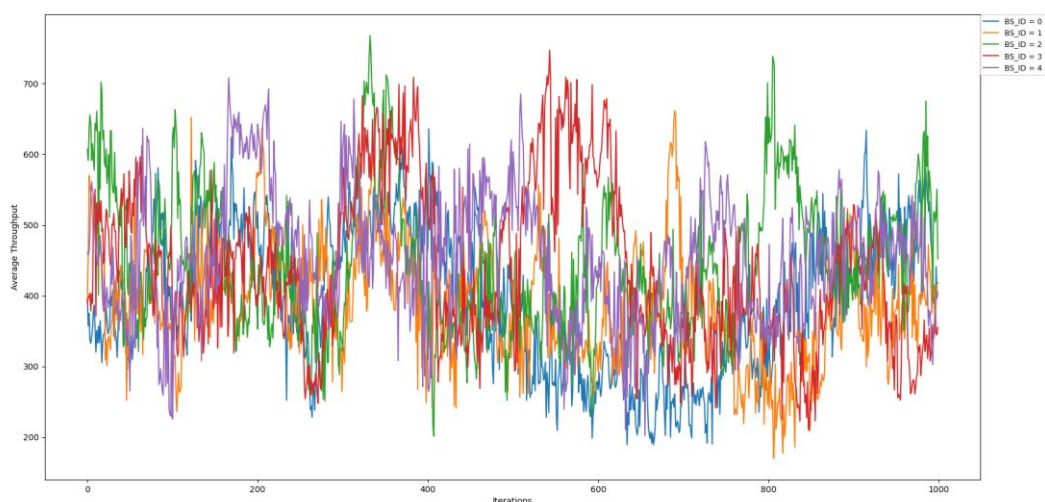


Σχήμα 22. Μέσο Throughput ανα BS

Iterations	BS_0	BS_1	BS_2	BS_3	BS_4
100	9.0	9.12	9.01	8.99	9.1
200	9.005	9.095	9.01	9.005	9.06
300	9.296	9.086	8.79	9.06	9.18
400	9.4975	9.24	8.86	9.05	9.1475
500	9.504	9.514	8.878	9.068	9.222
600	9.55	9.67	8.96	9.07	9.27
700	9.54	9.71	9.13	9.06	9.35
800	9.57	9.695	9.27	9.075	9.455
900	9.64	9.72	9.43	9.14	9.51
1000	9.652	9.763	9.378	9.256	9.634

Πίνακας 6. Μέσος αριθμός Συσχέτισης UE

Στην συνέχεια βλέπουμε τα αποτελέσματα χρησιμοποιώντας το Σχήμα 1. Μηχανισμός Handover που περιγράψαμε στο Κεφάλαιο 3.



Σχήμα 23. Μέσο Throughput με Handover.

Iterations	BS_0	BS_1	BS_2	BS_3	BS_4
100	14.55	16.46	14.08	17.94	14.48
200	14.51	17.1	13.83	18.1	15.135
300	14.39	16.98	13.70	17.32	15.17
400	14.27	17.32	13.6825	16.555	15.28
500	14.242	17.574	13.852	16.414	15.198
600	14.365	17.61	13.505	16.38	15.48
700	14.56	17.40	13.49	16.38	15.77
800	14.46	17.34	13.55	16.422	16.04875
900	14.384	17.29	13.52	16.424	16.824
1000	14.485	17.37	13.639	16.5	17.021

Πίνακας 7. Μέσος αριθμός Συσχέτισης με Handover

Μπορούμε να δούμε ότι παρόλο που με την χρήση του δικού μας μηχανισμού μερικοί χρήστες θα ανατεθούν σε BS που δεν τους προσφέρει το βέλτιστο Throughput (όπως είπαμε κάθε UE επιχειρεί να συνδεθεί στον BS όπου έχει το υψηλότερο SINR) καταφέρνουμε να αυξήσουμε τον μέσο ρυθμό μετάδοσης δεδομένων ανα BS. Αυτό οφείλεται στο ότι προτεραιότητα δίνουμε σε χρήστες με μικρές απαιτήσεις και ο μηχανισμός μας βοηθάει στο να οργανώσουμε τους πόρους του δικτύου με έναν πιο αποτελεσματικό τρόπο. Αυτό οδηγεί στο να εξυπηρετούμε χρήστες με μεγάλες απαιτήσεις που κανονικά θα έμεναν μη εξυπηρετημένοι. Έτσι το μέσο Throughput ανεβαίνει, ενώ το tradeoff είναι σχετικά μικρό αφού αυτοί οι χρήστες που δεν ανατίθενται στην καλύτερη επιλογή έχουν εξ αρχής μικρό Throughput. Ταυτόχρονα καταφέρνουμε να αυξήσουμε τον αριθμό των χρηστών που εξυπηρετούνται χωρίς κανένα αντάλλαγμα αφού οι χρήστες που γίνονται Handover σε άλλους BSs συνεχίζουν να εξυπηρετούνται απλά με χαμηλότερο ή ίσο Throughput.

5

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

5.1 Σύνοψη

Οι Στόχοι της εργασίας επιτεύχθηκαν με μεγάλη επιτυχία. Καταφέραμε να συγκρίνουμε διάφορους αλγόριθμους συσταδοποίησης ως προς την παραγωγή πραγματικών γεωγραφικών συντεταγμένων που θα αναπαριστούν Σταθμούς Βάσης και στην συνέχεια να οργανώσουμε το δίκτυο πάνω σε αυτούς τους. Η διεκπεραίωση ενός εργαλείου που να προσομοιώνει την κίνηση των χρηστών και τις μεταβαλλόμενες ιδιότητες που λαμβάνουν μέρος στην ενορχήστρωση ενός δικτύου, μας επέτρεψαν να μελετήσουμε πως αυτοί οι αλγόριθμοι συσταδοποίησης αποδίδουν ως προς τον αριθμό των χρηστών που μπορούν να εξυπηρετηθούν σε διάφορα στιγμιότυπα δικτύου, αλλά και πώς ο ρυθμός μετάδοσης δεδομένων συμπεριφέρεται σε κάθε ένα από αυτά τα στιγμιότυπα. Ακόμη είδαμε πως μπορούμε να επεκτείνουμε την απλή λειτουργία συσχέτισης UEs με BSs χρησιμοποιώντας μια τεχνική Handover δικής μας επινόησης, και με πάλι με την βοήθεια του εργαλείου προσομοίωσης είδαμε πως αυτή η τεχνική μπορεί να συνεισφέρει στην αύξηση του Throughput και του αριθμού εξυπηρετούμενων χρηστών, όπου ήταν και ο στόχος μας.

5.2 Μελλοντική Εργασία

Στο επίπεδο του προσομοιωτή, η μελλοντική δουλειά μπορεί να εστιάσει στην δημιουργία ενός Graphical User Interface που να επιτρέπει στον χρήστη να ορίζει τις παραμέτρους της προσομοίωσης, και κατ' επέκταση την αρχιτεκτονική του δικτύου, μέσω της εφαρμογής άμεσα. Όσο αφορά τα χαρακτηριστικά του δικτύου, μελλοντικές επεκτάσεις μπορούν να ενσωματώσουν την δυνατότητα για αρχιτεκτονικές τύπου HetNet όπου θα υποστηρίζονται και διαφορετικοί τύποι Σταθμών Βάσης πέρα των MacroCells, και την δυνατότητα για Downlink and Uplink Decoupling (DUDe) όπου τα UE θα έχουν την δυνατότητα να πραγματοποιούν συνδέσεις σε διαφορετικά BSs σε Downlink και Uplink. Αυτές οι αλλαγές είναι από μόνες τους είναι ικανές να αυξήσουν ραγδαία τον μέσο ρυθμό μετάδοσης δεδομένων, και θα είναι άξιο μελέτης να ερευνηθεί πώς αυτές οι τεχνολογίες σε συνδυασμό με τον δικό μας μηχανισμό θα επηρεάσουν το Throughput καθώς και τον αριθμό των συνδεδεμένων χρηστών.

ΒΙΒΛΙΟΓΡΑΦΙΑ- ΑΝΑΦΟΡΕΣ

- [1]. Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160.1 (2007): 3-24.
- [2]. Data Mining: Concepts and Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei
- [3]. Kaloxylos, Alexandros, Gavras, Anastasius, Camps Mur, Daniel, Ghorraishi, Mir, & Hrasnica, Halid. (2020). AI and ML – Enablers for Beyond 5G Networks. Zenodo. <https://doi.org/10.5281/zenodo.4299895>
- [4]. Yu, Kun & Tan, Lizhuang & Wu, Xiao-jin & Gai, Zhi-yong. (2019). Machine Learning Driven Network Routing. 705-712. 10.1109/ICSAI48974.2019.9010507.
- [5]. S. Ashtari, I. Zhou, M. Abolhasan, N. Shariati, J. Lipman, and W. Ni, "Knowledge-defined networking:
- [6]. S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [7]. H. Holma and A. Toskala, Eds., "LTE for UMTS: OFDMA and SC-FDMA based radio access", John Wiley & Sons, pp. 213-257, 2009
- [8]. Bouras, C., Kokkinos, V., Gkamas, A., Michos, E., Sina, I., Prokopiou, I., & Allayiotis, F. AN INTRODUCTION OF UPCOMING RADIO RESOURCE MANAGEMENT TECHNIQUES FOR 5G NETWORKS.
- [9]. "A BRIEF HISTORY OF MOBILE GENERATIONS AND SATELLITE WIRELESS COMMUNICATION SYSTEM ", *International Journal of Emerging Technologies and Innovative Research* (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.4, Issue 12, page no. pp1211-1216, December-2017
- [10]. R. Dilli, "Analysis of 5G Wireless Systems in FR1 and FR2 Frequency Bands," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 767-772, doi: 10.1109/ICIMIA48430.2020.9074973.
- [11]. F. Z. Yousaf, M. Bredel, S. Schaller and F. Schneider, "NFV and SDN—Key Technology Enablers for 5G Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2468-2478, Nov. 2017, doi: 10.1109/JSAC.2017.2760418.
- [12]. K. Saija, S. Nethi, S. Chaudhuri and R. M. Karthik, "A Machine Learning Approach for SNR Prediction in 5G Systems," 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2019, pp. 1-6, doi: 10.1109/ANTS47819.2019.9118097.
- [13]. Elgendi, Hesham & Maenpaa, Mikko & Levanen, Toni & Ihalainen, Tero & Nielsen, Sari & Valkama, Mikko. (2019). Interference Measurement Methods in 5G NR: Principles and Performance. 233-238. 10.1109/ISWCS.2019.8877215.
- [14]. L. Le, B. P. Lin, L. Tung and D. Sinh, "SDN/NFV, Machine Learning, and Big Data Driven Network Slicing for 5G," 2018 IEEE 5G World Forum (5GWF), 2018, pp. 20-25, doi: 10.1109/5GWF.2018.8516953.
- [15]. A. K. Thyagarajan, P. Balasubramanian, V. D and K. M, "SNR-CQI Mapping for 5G Downlink Network," 2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), 2021, pp. 173-177, doi: 10.1109/APWiMob51111.2021.9435258.

- [16]. 3GPP: 3GPP TS 38.104, v15.1.0, Base Station (BS) radio transmission and reception. Release 15 (2018)
- [17]. 3GPP TS 38.215: 5G NR; Physical layer measurements
- [18]. 3GPP TS 38.306 v16.3 - 4.1.2 Supported max data rate for DL/U
- [19]. Alcatel-Lucent. "Application of Machine-Learning Based Prediction Techniques in Wireless Networks." (2014).
- [20]. Bouras, Christos & Kalogeropoulos, Rafail. (2021). User Allocation in 5G Networks Using Machine Learning Methods for Clustering. 10.1007/978-3-030-75100-5_2.
- [21]. Kamal MA, Raza HW, Alam MM, Su'ud MM, Sajak ABAB. Resource Allocation Schemes for 5G Network: A Systematic Review. *Sensors (Basel)*. 2021;21(19):6588. Published 2021 Oct 2. doi:10.3390/s21196588
- [22]. Klisara, Sara & Goran, N. & Avdagić-Golub, Elma. (2021). Rough estimation of cell numbers in 5G networks using simple mathematical calculations. *Science, Engineering and Technology*. 1. 1-7. 10.54327/set2021/v1.i2.15.
- [23]. J. Lee and Y. Yoo, "Handover cell selection using user mobility information in a 5G SDN-based network," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), 2017, pp. 697-702, doi: 10.1109/ICUFN.2017.7993880.
- [24]. Kosmides, Pavlos & Adamopoulou, Evgenia & Demestichas, Konstantinos & Anagnostou, Miltiades & Rouskas, Angelos. (2015). On Intelligent Base Station Activation for Next Generation Wireless Networks. *Procedia Computer Science*. 63. 82-88. 10.1016/j.procs.2015.08.315.
- [25]. R. Nordin, M. H. Alsharif, G. Woodward and N. Pau, "Dynamic resource block for base station grouping in macro-diversity systems for LTE public safety networks," *2014 IEEE 2nd International Symposium on Telecommunication Technologies (ISTT)*, 2014, pp. 245-250, doi: 10.1109/ISTT.2014.7238213.
- [26]. Bouras, Christos & Kollia, Anastasia & Papazois, Andreas. (2018). Exploring SDN & NFV in 5G Using ONOS & POX Controllers. *International Journal of Interdisciplinary Telecommunications and Networking*. 10. 46-60. 10.4018/IJITN.2018100103.
- [27]. Bouras, Christos & Kalogeropoulos, Rafail. (2021). User Allocation in 5G Networks Using Machine Learning Methods for Clustering. 10.1007/978-3-030-75100-5_2.
- [28]. Bouras, Christos & Kalogeropoulos, Rafail. (2021). User Allocation in 5G Networks Using Machine Learning Methods for Clustering. 10.1007/978-3-030-75100-5_2.
- [29]. Bouras, Christos & Diles, Georgios & Kalogeropoulos, Rafail. (2020). An Energy Efficient Mechanism for Downlink and Uplink Decoupling in 5G Networks. 10.1007/978-3-030-33506-9_22.
- [30]. Bouras, Christos, and Rafail Kalogeropoulos. "A QoS driven adaptive mechanism for downlink and uplink decoupling in 5G." *Internet of Things* 11 (2020): 100217.
- [31]. Bouras, Christos & Kokkinos, Vasileios & Michos, Evangelos. (2020). A Resource-Efficient Approach on User Association in 5G Networks Using Downlink and Uplink Decoupling. *International Journal of Wireless Networks and Broadband Technologies*. 9. 43-59. 10.4018/IJWNBT.2020070103.
- [32]. Bouras, Christos, Vasileios Kokkinos, and Evangelos Michos. "Efficient 5G Network Decoupling Using Dynamic Modulation and Coding Scheme Selection." *International Conference on Broadband and Wireless Computing, Communication and Applications*. Springer, Cham, 2019.
- [33]. BANNOUR, AHMED & Harbaoui, Ahmed & Alsolami, Fawaz. (2021). Connected Objects Geo-Localization Based on SS-RSRP of 5G Networks. *Electronics*. 10. 10.3390/electronics10222750.

Ιστοσελίδες

- [34]. <https://en.wikipedia.org/wiki/BIRCH>
- [35]. https://en.wikipedia.org/wiki/Reinforcement_learning
- [36]. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
- [37]. [https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))
- [38]. <https://www.rfwireless-world.com/>
- [39]. <http://howltestuffworks.blogspot.com/search/label/5G>
- [40]. https://www.sharetechnote.com/html/5G/5G_FR_Bandwidth.html
- [41]. <https://www.techplayon.com>

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ ΣΥΓΓΡΑΦΕΑ



Ο Μάριος Χαραλαμπίδης είναι τελειόφητος προπτυχιακός φοιτητής στο τμήμα Μηχανικών Η/Υ και Πληροφορικής της Πάτρας. Αποφοίτησε από το Λύκειο Κύκκου Πάφου το 2014. Το ίδιο έτος εντάχθηκε στην Εθνική Φρουρά όπου υπηρέτησε την στρατιωτική του θητεία ως Δόκιμος Έφεδρος Αξιωματικός μέχρι το 2016. Από το 2019 αποτελεί μέλος του εργαστηρίου Κατανεμημένων Συστημάτων και Τηλεματικής στο οποίο ασχολείται με την ανάλυση δεδομένων σε δίκτυα κινητής τηλεφωνίας. Έχει εργασθεί σε έργα του Ειδικού Λογαριασμού Κονδυλίων Έρευνας όπου ανέπτυξε γνώσεις πάνω σε σχεδίαση και ανάπτυξη ιστοσελίδων και σχεδίαση και διαχείριση Βάσεων Δεδομένων