**Δομή Παρουσίασης PowerPoint**

1. **Εισαγωγική Διαφάνεια**: **Τι είναι το 5G;**
   * Βασικά χαρακτηριστικά του 5G (υψηλότερες ταχύτητες, μειωμένη καθυστέρηση, αυξημένη συνδεσιμότητα)
   * Πιθανές εφαρμογές του 5G (Ανάλυση Δεδομένων, IoT, αυτόνομα οχήματα, εξυπηρέτηση πελατών)
2. **Πώς το 5G συνδέεται με τα Δεδομένα μας**
   * Επισκόπηση των δεδομένων και των κύριων μεταβλητών
   * Περιγραφή της σχέσης κάθε μεταβλητής με τις δυνατότητες του 5G
3. **Περιγραφική Στατιστική Ανάλυση**
   * Γραφήματα και στατιστικές που δείχνουν κρίσιμες κατανομές και τάσεις
   * Σχολιασμός σχετικά με τη σημασία κάθε αποτελέσματος
4. **Εισαγωγή στη Μηχανική Μάθηση**
   * Πώς η Μηχανική Μάθηση μπορεί να εφαρμοστεί στα δεδομένα
   * Εισαγωγή της τεχνικής XGBoost και της εφαρμογής της στα δεδομένα μας
5. **Συμπεράσματα και Μελλοντικά Βήματα**
   * Σύνοψη των κυρίων ευρημάτων
   * Προτάσεις για την περαιτέρω ανάλυση και χρήση του 5G

………………………………………………………………………………………………………………Το 5G αντιπροσωπεύει τη νέα γενιά τηλεπικοινωνιακών τεχνολογιών, που προσφέρει αναβαθμισμένες δυνατότητες επικοινωνίας και σύνδεσης σε σύγκριση με τις προηγούμενες γενιές. Το κύριο χαρακτηριστικό του 5G είναι η υψηλή ταχύτητα μετάδοσης δεδομένων, που επιτρέπει τη λήψη και τη μετάδοση μεγάλου όγκου πληροφοριών με εκπληκτική ταχύτητα. Επιπλέον, το 5G χαρακτηρίζεται από χαμηλότερη καθυστέρηση στη μετάδοση δεδομένων, γεγονός που οδηγεί σε αποτελεσματικότερη και αμεσότερη επικοινωνία. Τέλος, μια από τις κύριες επιδιώξεις του 5G είναι η επίτευξη μεγαλύτερης συνδεσιμότητας, που θα επιτρέψει στις συσκευές να είναι συνδεδεμένες μεταξύ τους με μεγαλύτερη αξιοπιστία και αποτελεσματικότητα. Αυτά τα χαρακτηριστικά καθιστούν το 5G μια εξαιρετικά ελκυστική τεχνολογία με πληθώρα προοπτικών και δυνατοτήτων για τη μελλοντική ανάπτυξη και εξέλιξη των τηλεπικοινωνιακών δικτύων.

Το 5G έχει σημαντική επίδραση και στην ίδια την ανάλυση δεδομένων και τις τεχνολογίες συλλογής δεδομένων λόγω των χαρακτηριστικών του όπως περιεγράφηκαν παραπάνω:

1. **Υψηλή Ταχύτητα και Χωρητικότητα:** Το 5G παρέχει υψηλότερες ταχύτητες μετάδοσης δεδομένων και μεγαλύτερη χωρητικότητα σε σύγκριση με τις προηγούμενες γενιές δικτύων. Αυτό επιτρέπει την ταχύτερη συλλογή, μετάδοση και επεξεργασία μεγάλων όγκων δεδομένων.
2. **Χαμηλή Καθυστέρηση:** Η χαμηλή καθυστέρηση στη μετάδοση δεδομένων που προσφέρει το 5G επιτρέπει την πραγματοποίηση πραγματικού χρόνου ανάλυσης δεδομένων και ανταπόκρισης σε πραγματικό χρόνο σε διάφορες εφαρμογές, όπως τα αυτόνομα οχήματα, οι υπηρεσίες υγείας και η βιομηχανία.
3. **Μεγαλύτερη Συνδεσιμότητα:** Η αυξημένη συνδεσιμότητα που προσφέρει το 5G επιτρέπει τη σύνδεση μεγαλύτερου αριθμού συσκευών και αισθητήρων σε ένα δίκτυο, δημιουργώντας ένα πλήθος νέων πηγών δεδομένων.
4. **Καινοτόμες Εφαρμογές:** Το 5G επιτρέπει τη δημιουργία καινοτόμων εφαρμογών που βασίζονται στην ανάλυση δεδομένων, όπως η επαυξημένη πραγματικότητα, οι ευφυείς πόλεις, οι ιατρικές τηλεματικές υπηρεσίες και η βιομηχανική αυτοματοποίηση.

Οι τεχνολογίες συλλογής δεδομένων είναι αναγκαίες για τη συλλογή, την αποθήκευση και την επεξεργασία των δεδομένων που παράγονται από τα δίκτυα 5G και τις συνδεδεμένες συσκευές. Αυτές οι τεχνολογίες περιλαμβάνουν αισθητήρες IoT, κεντρικές μονάδες επεξεργασίας δεδομένων (CPU), συστήματα αποθήκευσης δεδομένων (όπως cloud και edge computing) και λογισμικό ανάλυσης δεδομένων. Η εξέλιξη αυτών των τεχνολογιών συμβάλλει στην ανάπτυξη καινοτόμων λύσεων ανάλυσης δεδομένων που εκμεταλλεύονται τα πλεονεκτήματα του 5G.

………………………………………………………………………………………………………………

Στο δεύτερο σκέλος της παρουσίασής μας εξερευνούμε τα δεδομένα που διαθέτουμε και τις μεταβλητές που περιλαμβάνονται σε αυτά.

Περιγραφή των Δεδομένων:

Τα δεδομένα μας αποτελούν ένα σύνολο που περιέχει πληροφορίες σχετικά με την ποιότητα της εξυπηρέτησης σε ένα δίκτυο 5G. Κάθε εγγραφή αντιστοιχεί σε έναν χρήστη και περιλαμβάνει πληροφορίες σχετικά με το είδος της εφαρμογής που χρησιμοποιείται, την ισχύ του σήματος, το χρόνο καθυστέρησης στη μετάδοση δεδομένων, καθώς και πληροφορίες σχετικά με την απαιτούμενη και τη διατεθείσα εύρος ζώνης, καθώς και το ποσοστό των πόρων που έχουν εκχωρηθεί στην εφαρμογή.

Ας δούμε πώς η κάθε μεταβλητή σχετίζεται με τις δυνατότητες του 5G:

1. **Application Type (Τύπος Εφαρμογής):** Αυτή η μεταβλητή περιγράφει το είδος της εφαρμογής που χρησιμοποιείται στο δίκτυο. Το 5G επιτρέπει τη λειτουργία ποικίλων εφαρμογών με διαφορετικές απαιτήσεις σε ταχύτητα, χωρητικότητα και χρόνο απόκρισης.
2. **Signal Strength (Ισχύς Σήματος):** Η ισχύς του σήματος επηρεάζει την ποιότητα της επικοινωνίας σε ένα δίκτυο 5G. Ένα ισχυρό σήμα εξασφαλίζει σταθερή και γρήγορη σύνδεση, ενώ ένα ασθενές σήμα μπορεί να οδηγήσει σε απώλεια σύνδεσης ή χαμηλή ταχύτητα μετάδοσης δεδομένων.
3. **Latency (Καθυστέρηση):** Η καθυστέρηση στη μετάδοση δεδομένων είναι κρίσιμη για πολλές εφαρμογές, όπως τα αυτόνομα οχήματα και οι εικονικές πραγματικότητες. Το 5G προσφέρει χαμηλή καθυστέρηση, επιτρέποντας την πραγματοποίηση ανταπόκρισης σε πραγματικό χρόνο.
4. **Required Bandwidth (Απαιτούμενη Εύρος Ζώνης):** Η απαίτηση εύρους ζώνης καθορίζει την απόδοση της εφαρμογής. Το 5G παρέχει μεγαλύτερη χωρητικότητα στο δίκτυο, επιτρέποντας τη μετάδοση μεγάλων όγκων δεδομένων με υψηλές ταχύτητες.
5. **Allocated Bandwidth (Εκχωρημένο Εύρος Ζώνης):** Η εκχώρηση εύρους ζώνης σε μια εφαρμογή επηρεάζει την ταχύτητα και την απόδοση της μετάδοσης δεδομένων σε αυτήν.
6. **Resource Allocation (Κατανομή Πόρων):** Η κατανομή πόρων σε μια εφαρμογή είναι σημαντική για την εξασφάλιση της αποτελεσματικής λειτουργίας της σε ένα δίκτυο 5G. Η ορθή κατανομή πόρων επιτρέπει τη βέλτιστη χρήση τους και την εξασφάλιση της απόδοσης της εφαρμογής.

Συνολικά, οι μεταβλητές μας αντικατοπτρίζουν την ποιότητα της υπηρεσίας σε ένα δίκτυο 5G και τις δυνατότητες που παρέχει για τη συλλογή, μετάδοση και επεξεργασία δεδομένων με υψηλές ταχύτητες, χαμηλή καθυστέρηση και αποτελεσματική χρήση πόρων. Η ανάλυση και η κατανόηση αυτών των δεδομένων μας επιτρέπει να λάβουμε σημαντικά συμπεράσματα και να προβλέψουμε την απόδοση της εφαρμογής σε διάφορες συνθήκες.

………………………………………………………………………………………………………………

**Παρατηρήσεις από τα Ιστογράμματα:**

* **resource\_allocation**: Η κατανομή είναι ασύμμετρη δεξιά, με την πλειοψηφία των τιμών να συγκεντρώνεται γύρω στο 70%. Αυτό υποδηλώνει ότι οι περισσότεροι χρήστες λαμβάνουν παρόμοια κατανομή πόρων. Η ασύμμετρη δεξιά κατανομή υποδηλώνει ότι λιγότερες περιπτώσεις χρήζουν πολύ υψηλής κατανομής πόρων. Η κεντρική τάση των τιμών στο 70% μπορεί να υποδηλώνει μία κοινή βάση για τις ανάγκες των περισσότερων εφαρμογών. Αυτό μπορεί να είναι χρήσιμο στην πρόβλεψη της latency, ειδικά σε συνδυασμό με τις απαιτήσεις bandwidth.
* **required\_bandwidth (Mbps)**: Η κατανομή είναι ασύμμετρη δεξιά, με λίγες εφαρμογές να απαιτούν μεγάλο εύρος ζώνης.
* **allocated\_bandwidth (Mbps)**: Η κατανομή είναι παρόμοια με το required\_bandwidth, υποδηλώνοντας μια συσχέτιση μεταξύ απαιτούμενου και εκχωρημένου εύρους ζώνης.

Η ασυμμετρία δεξιά και η παρόμοια κατανομή ανάμεσα σε αυτές τις δύο μεταβλητές υποδηλώνει πως το υψηλότερο required bandwidth μπορεί να οδηγεί στην αντίστοιχη αύξηση του allocated bandwidth. Αυτό θα μπορούσε να έχει απτές επιπτώσεις στην latency.

* **latency (msec)**: Η κατανομή είναι ελαφρώς ασύμμετρη δεξιά, με την πλειοψηφία των τιμών να κυμαίνεται μεταξύ 0 και 40 ms. Υπάρχουν όμως και κάποιες περιπτώσεις με μεγαλύτερη καθυστέρηση. Η κατανομή της latency υποδηλώνει ότι η πλειοψηφία των ερωτημάτων επεξεργάζεται με σχετικά χαμηλή καθυστέρηση, αλλά υπάρχουν και κάποιες ακραίες περιπτώσεις. Αυτές οι ακραίες τιμές θα μπορούσαν να είναι σημαντικές για την πρόβλεψη και μπορεί να απαιτούν ειδική ανάλυση ή μεταχείριση.
* **signal\_strength (dBm)**: Η κατανομή μοιάζει κανονική (Gaussian) με τις περισσότερες τιμές να συγκεντρώνονται γύρω στο -80 dBm. Η κανονική κατανομή της ισχύος του σήματος υποδηλώνει ότι οι τιμές του σήματος είναι σχετικά ομοιόμορφα διασπαρμένες. Αυτό προσδίδει ευκολία στην ερμηνεία και μπορεί να είναι ένας σταθερός προβλεπτικός παράγοντας για την καθυστέρηση ή άλλες μεταβλητές.

**Παρατηρήσεις από το Γράφημα Πίτας:**

* Οι τύποι εφαρμογών με τη μεγαλύτερη συχνότητα είναι "Web Browsing", "Background Download" και "Video Call".
* Οι τύποι εφαρμογών "IoT Temperature" και "Emergency Service" έχουν τη χαμηλότερη συχνότητα.
* Η κατανομή των τύπων εφαρμογών μπορεί να δώσει επιπλέον πληροφορίες για την απόδοση δικτύου, καθώς διαφορετικοί τύποι εφαρμογών έχουν διαφορετικές ανάγκες για bandwidth και latency. Για παράδειγμα, οι video calls απαιτούν υψηλότερη ποιότητα σήματος και ευρύτερο bandwidth σε σύγκριση με το web browsing ή τις background downloads.

**Συμπεράσματα:**

* Η κατανόηση της κατανομής και των συσχετίσεων μεταξύ των μεταβλητών είναι κρίσιμη για την επιλογή και την βελτιστοποίηση του μοντέλου μηχανικής μάθησης.
* Είναι πιθανό ο τύπος της εφαρμογής, η ισχύς του σήματος και το εκχωρημένο εύρος ζώνης να είναι σημαντικοί προβλεπτικοί παράγοντες για την καθυστέρηση.
* Περαιτέρω ανάλυση, όπως η εξέταση συσχετίσεων και η εφαρμογή τεχνικών μείωσης διαστάσεων, μπορεί να βελτιώσει την απόδοση του μοντέλου.

Συμπεράσματα και Προτάσεις

Η διεξοδική ανάλυση των μεταβλητών και η κατανόηση των συσχετίσεών τους είναι κρίσιμα σημεία για την επιλογή του κατάλληλου μοντέλου μηχανικής μάθησης. Η εξέταση συσχετίσεων μέσω scatter plots ή correlation matrices και η χρήση τεχνικών μείωσης διαστάσεων όπως PCA μπορεί να βοηθήσει στην αποφυγή πολυπλοκότητας και στη βελτίωση της προβλεπτικής ακρίβειας του τελικού μοντέλου.

………………………………………………………………………………………………………………

Η Μηχανική Μάθηση μπορεί να εφαρμοστεί στα δεδομένα μας με σκοπό την πρόβλεψη της καθυστέρησης (Latency) στο δίκτυο 5G με βάση τις διαφορετικές μεταβλητές που διαθέτουμε. Μέσω μηχανικών μοντέλων μπορούμε να εξάγουμε πολύπλοκες σχέσεις από τα δεδομένα και να προβλέψουμε τις τιμές της καθυστέρησης για νέες καταστάσεις.

Το XGBoost (Extreme Gradient Boosting) είναι μια προηγμένη μέθοδος μηχανικής μάθησης που χρησιμοποιείται ευρέως για προβλήματα παλινδρόμησης και ταξινόμησης. Αποτελεί έναν από τους πιο αποδοτικούς αλγορίθμους μηχανικής μάθησης και είναι εξαιρετικά δημοφιλής λόγω της ικανότητάς του να παρέχει υψηλή ακρίβεια και να αντιμετωπίζει αποτελεσματικά την υπερεκπαίδευση.

Για την εφαρμογή του XGBoost στα δεδομένα μας, θα δημιουργήσουμε ένα μοντέλο παλινδρόμησης όπου η μεταβλητή target θα είναι η καθυστέρηση (Latency), ενώ οι υπόλοιπες μεταβλητές θα είναι οι προβλεπόμενοι predictors. Στη συνέχεια, θα εκπαιδεύσουμε το μοντέλο μας χρησιμοποιώντας τον XGBoost regressor, ρυθμίζοντας τα υπερπαραμέτρους του αλγορίθμου με την τεχνική του cross-validation, προκειμένου να επιτύχουμε την καλύτερη δυνατή ακρίβεια και γενικευμένη απόδοση του μοντέλου μας.

Συγκεκριμένα η εκπαίδευση του μοντέλου XGBoost ακολουθεί τα παρακάτω βήματα και με αυτόν τον τρόπο θα προχωρήσουμε την εργασία μας:

1. **Ορισμός των Παραμέτρων Του Μοντέλου**: Αυτό περιλαμβάνει την επιλογή των υπερπαραμέτρων του μοντέλου, όπως οι μέγιστες βάθος των δέντρων αποφάσεων, η ταχύτητα μάθησης (learning rate), και ο αριθμός των δέντρων στο σύνολο (number of trees).
2. **Εκπαίδευση του Μοντέλου**: Κατά την εκπαίδευση, το μοντέλο προσαρμόζεται στα δεδομένα εκπαίδευσης με σκοπό να μάθει τη σχέση μεταξύ των προβλεπόμενων παραμέτρων και της μεταβλητής "latency".
3. **Προσαρμογή Παραμέτρων Μοντέλου**: Κατά τη διάρκεια της εκπαίδευσης, οι υπερπαράμετροι του μοντέλου μπορεί να προσαρμοστούν προκειμένου να βελτιστοποιηθεί η απόδοσή του.
4. **Αξιολόγηση του Μοντέλου**: Μετά την εκπαίδευση, το μοντέλο αξιολογείται χρησιμοποιώντας τα δεδομένα ελέγχου. Αυτό γίνεται για να εκτιμηθεί η απόδοσή του μοντέλου σε ανεξάρτητα δεδομένα.
5. **Πρόβλεψη**: Το εκπαιδευμένο μοντέλο χρησιμοποιείται για να προβλέψει τις τιμές της μεταβλητής "latency" για νέα δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση.

……………………………………………………………………………………………………………… **Συμπεράσματα και Μελλοντικά Βήματα**

**Σύνοψη των κυρίων ευρημάτων:**

Η ανάλυση των δεδομένων δικτύου 5G που διεξήχθη έδειξε σημαντικές παρατηρήσεις αναφορικά με την κατανομή και τις ανάγκες των πόρων (resource allocation), την ισχύ του σήματος, και την καθυστέρηση (latency). Ειδικότερα:

1. **Resource Allocation**: Η κατανομή των πόρων εμφανίζεται να είναι ασύμμετρη με δεξιά κλίση, υποδεικνύοντας ότι η μεγαλύτερη ανάγκη για πόρους εστιάζεται σε λιγότερες αλλά πιο απαιτητικές εφαρμογές.
2. **Bandwidth Requirements**: Υπάρχει στενή συσχέτιση ανάμεσα στο απαιτούμενο και το πραγματικά διαθέσιμο bandwidth, υποδηλώνοντας έναν επαρκή συγχρονισμό μεταξύ αναγκών και πόρων.
3. **Latency**: Παρότι η καθυστέρηση είναι γενικά χαμηλή, οι ακραίες τιμές καθυστέρησης χρήζουν περαιτέρω ανάλυσης για την κατανόηση των παραγόντων που τις επηρεάζουν.
4. **Signal Strength**: Η σχετικά κανονική κατανομή της ισχύος του σήματος δείχνει ομοιογένεια στη διανομή της ποιότητας του δικτύου.

**Προτάσεις για την περαιτέρω ανάλυση και χρήση του 5G:**

1. **Βελτιστοποίηση Resource Allocation**: Αναπτύξτε μοντέλα πρόβλεψης που θα εντοπίζουν τις πιο απαιτητικές εφαρμογές και θα προτείνουν δυναμική εκχώρηση πόρων για τη μείωση της καθυστέρησης και την αυξημένη απόδοση.
2. **Ανάλυση Καθυστέρησης**: Εξετάστε λεπτομερώς τις αιτίες των ακραίων τιμών της καθυστέρησης μέσω βαθύτερης διερεύνησης των αιτιών και των συσχετίσεων με άλλες μεταβλητές.
3. **Μοντελοποίηση Πρόβλεψης Ισχύος Σήματος**: Αναπτύξτε προβλεπτικά μοντέλα για την ισχύ του σήματος βασιζόμενα σε παράγοντες όπως η τοποθεσία του χρήστη και το περιβάλλον για να βελτιώσετε τη συνολική εμπειρία του χρήστη.
4. **Τεχνικές Εξορυκτικής Δεδομένων**: Εφαρμόστε τεχνικές όπως machine learning και data mining για να ανακαλύψετε νέα μοτίβα και συσχετίσεις στα δεδομένα που μπορεί να συμβάλουν στη βελτίωση των στρατηγικών διαχείρισης του δικτύου.
5. **Εφαρμογή Διαχείρισης Κατανάλωσης Ενέργειας**: Αναλύστε την κατανάλωση ενέργειας και αναπτύξτε μέθοδους για τη βελτιστοποίηση της ενεργειακής απόδοσης του δικτύου, μειώνοντας τον περιβαλλοντικό αντίκτυπο και το κόστος.

Με αυτές τις προτάσεις, ελπίζουμε να παρέχουμε μια βάση για περαιτέρω βελτίωση και αξιοποίηση των δυνατοτήτων της τεχνολογίας 5G, με στόχο την καλύτερη εξυπηρέτηση των χρηστών και την αποδοτικότερη χρήση των τεχνολογικών πόρων.

………………………………………………………………………………………………………………

Αυτό το εισαγωγικό PowerPoint μπορεί να χρησιμοποιηθεί ως αρχική παρουσίαση προτού μπούμε σε λεπτομερή ανάλυση και μοντελοποίηση των δεδομένων μας. Εάν έχετε οποιεσδήποτε πρόσθετες απαιτήσεις ή προτάσεις, παρακαλώ ενημερώστε μας!