Ας ξεκινήσουμε την ανάλυση των δεδομένων χρησιμοποιώντας τον XGBoost regressor. Αρχικά, θα πραγματοποιήσουμε την προεπεξεργασία των δεδομένων σύμφωνα με τις προδιαγραφές που παρέχετε και στη συνέχεια θα διεξάγουμε το προκαταρκτικό στατιστικό στάδιο.

Προεπεξεργασία Δεδομένων:

1. Μετατροπή της ποιοτικής μεταβλητής "application type" σε διακριτά διανύσματα (dummies).
2. Κανονικοποίηση (min-max scaling) όλων των μεταβλητών εκτός από τη μεταβλητή που είναι target (latency).
3. Αντικατάσταση των missing values με 0.

Μετά την προεπεξεργασία των δεδομένων, μπορούμε να προχωρήσουμε στο προκαταρκτικό στατιστικό στάδιο.

Προκαταρκτική Ανάλυση Δεδομένων:

Ποιοτική Μεταβλητή "Application Type": Ας ξεκινήσουμε με μια πίτα που θα δείχνει το ποσοστό εμφάνισης κάθε είδους εφαρμογής.

Ποσοτικές Μεταβλητές:

Signal Strength (dBm), Latency (msec), Required Bandwidth (Mbps), Allocated Bandwidth (Mbps), Resource Allocation (%): Για κάθε μία από αυτές τις μεταβλητές, θα δημιουργήσουμε ένα ιστόγραμμα για να δούμε την κατανομή των δεδομένων. Στο ίδιο διάγραμμα θα περιλαμβάνουμε και τη μέση τιμή (μ) και το πρότυπο απόκλισης (σ) για κάθε μεταβλητή.

Συσχέτιση μεταβλητών με το Target (Latency): Θα χρησιμοποιήσουμε scatterplots με linear trendlines για να εξετάσουμε τυχόν σχέσεις μεταξύ των προβλέπτων μεταβλητών και του target (latency).

Θα ξεκινήσουμε την υλοποίηση αυτών των βημάτων σε Python με τη χρήση της βιβλιοθηκών.

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

Στο τρίτο σκέλος της παρουσίασής μας εξερευνούμε τη σχέση των δεδομένων μας με τα χαρακτηριστικά του 5G και πώς αυτά αντικατοπτρίζουν την ποιότητα της υπηρεσίας σε ένα δίκτυο 5G.

Οι μεταβλητές που έχουμε στη διάθεσή μας αποτελούν ουσιώδη στοιχεία για τη μέτρηση και την αξιολόγηση της απόδοσης της εφαρμογής σε ένα δίκτυο 5G. Η μέτρηση της ισχύος του σήματος, η καθυστέρηση στη μετάδοση δεδομένων και η εκχώρηση εύρους ζώνης είναι χαρακτηριστικά που άμεσα επηρεάζουν την εμπειρία του χρήστη σε ένα δίκτυο 5G.

Η ισχύς του σήματος είναι κρίσιμη για την επικοινωνία μεταξύ των συσκευών και την ποιότητα των υπηρεσιών. Μεγαλύτερη ισχύς σήματος σημαίνει πιο σταθερή και αξιόπιστη σύνδεση, ενώ χαμηλότερη ισχύς μπορεί να οδηγήσει σε απώλεια σήματος και μειωμένη απόδοση.

Η καθυστέρηση στη μετάδοση δεδομένων επηρεάζει την απόκριση του δικτύου και την απόδοση της εφαρμογής. Χαμηλότερη καθυστέρηση σημαίνει ότι οι χρήστες λαμβάνουν τις πληροφορίες πιο γρήγορα και με μικρότερο χρόνο αναμονής.

Το εύρος ζώνης είναι επίσης κρίσιμο για την απόδοση της εφαρμογής, καθώς καθορίζει την ποσότητα δεδομένων που μπορεί να μεταδοθεί σε ένα χρονικό διάστημα. Η εκχώρηση εύρους ζώνης στην εφαρμογή επηρεάζει την ταχύτητα και την ποιότητα της μετάδοσης δεδομένων.

Με την ανάλυση αυτών των χαρακτηριστικών και τη σύγκρισή τους με τις απαιτήσεις και τις αναμονές των χρηστών, μπορούμε να κατανοήσουμε πώς τα δεδομένα μας αντανακλούν την ποιότητα της υπηρεσίας σε ένα δίκτυο 5G και να προβλέψουμε την απόδοση της εφαρμογής σε διάφορες συνθήκες λειτουργίας.