

Αποτελεσματικότητα των Δύο Μοντέλων

1. Linear Regression:

- Το Linear Regression είναι ένα απλό γραμμικό μοντέλο και ενδείκνυται όταν η σχέση μεταξύ των χαρακτηριστικών και της μεταβλητής-στόχου είναι γραμμική.
- Σε ένα περιβάλλον όπου οι παράγοντες που επηρεάζουν την παραμονή στο νοσοκομείο είναι σχετικά απλοί και γραμμικοί, το Linear Regression μπορεί να δώσει ικανοποιητικά αποτελέσματα.
- Ωστόσο, αν η σχέση είναι πιο περίπλοκη ή μη γραμμική, η απόδοσή του πιθανόν να είναι χαμηλότερη.

2. Random Forest:

- Το Random Forest είναι ένα πιο πολύπλοκο, μη γραμμικό μοντέλο που μπορεί να συλλάβει πολύπλοκες σχέσεις μεταξύ των χαρακτηριστικών.
- Ενδείκνυται όταν υπάρχουν πολλές αλληλεπιδράσεις ή μη γραμμικές σχέσεις, όπως μπορεί να συμβαίνει σε ένα κλινικό περιβάλλον όπου πολλοί παράγοντες (όπως η κατάσταση του ασθενούς, η κλινική, οι διαθέσιμοι πόροι) αλληλεπιδρούν με πολυπλοκότητα.
- Αν το Random Forest έχει πολύ καλύτερη απόδοση από το Linear Regression, αυτό υποδηλώνει ότι οι παράγοντες που επηρεάζουν την παραμονή είναι μη γραμμικοί.

Επίδραση της Διαφοράς Αποτελεσματικότητας στην Πράξη

• Ακριβέστερο Μοντέλο:

- Ένα πιο ακριβές μοντέλο μπορεί να βελτιώσει τη διαχείριση των κλινικών πόρων. Για παράδειγμα, αν γνωρίζουμε με ακρίβεια τη διάρκεια παραμονής των ασθενών, η κλινική μπορεί να διαχειριστεί καλύτερα τις κλίνες, το προσωπικό και τα φάρμακα.
- Ένα ανακριβές μοντέλο, από την άλλη, μπορεί να οδηγήσει σε ελλείψεις ή υπερβολές, επηρεάζοντας την ποιότητα φροντίδας και την ασφάλεια των ασθενών.

• Διαφορά στα Δύο Μοντέλα:

- Αν η διαφορά στην αποτελεσματικότητα των δύο μοντέλων είναι μεγάλη, αυτό σημαίνει ότι η χρήση ενός πιο απλού μοντέλου (όπως το Linear Regression) μπορεί να οδηγήσει σε ανακριβείς προβλέψεις, κάτι που στην πράξη θα μπορούσε να έχει σημαντικές συνέπειες, όπως ακριβότερη φροντίδα ή περιορισμούς στη διαχείριση κλινικών πόρων.

Πειραματισμός Χωρίς τη Μεταβλητή “facid”

• Αφαίρεση της “facid”:

- Αν αφαιρέσουμε τη μεταβλητή “facid”, η απόδοση των μοντέλων ενδέχεται να μειωθεί, καθώς η “facid” μπορεί να περιέχει κρίσιμες πληροφορίες για τις διαφορές μεταξύ των κλινικών.
- Οι κλινικές ενδέχεται να έχουν διαφορετικές πολιτικές, πόρους και πρωτόκολλα, τα οποία επηρεάζουν τη διάρκεια παραμονής των ασθενών.

• Ανάλυση Αποτελεσμάτων:

- Εάν παρατηρήσουμε σημαντική πτώση της ακρίβειας χωρίς τη μεταβλητή “facid”, αυτό σημαίνει ότι η κλινική καταγωγή του ασθενούς (το "facility ID") είναι σημαντικός παράγοντας για την πρόβλεψη της διάρκειας παραμονής.
- Η απουσία της μεταβλητής θα μπορούσε να οδηγήσει σε γενίκευση των προβλέψεων, χωρίς να λαμβάνονται υπόψη οι ιδιαιτερότητες της κάθε κλινικής, γεγονός που μπορεί να οδηγήσει σε λανθασμένες αποφάσεις στην πράξη.

Συμπεράσματα

- Θα ήθελα να ενημερώσω ότι αντιμετώπισα κάποια δυσκολία στο κομμάτι με τον Normalizer και συγκεκριμένα τον Normalizer (Apply) και προσπάθησα να κάνω από εκεί και πέρα ότι μπορούσα!
- Ένα πιο ακριβές μοντέλο, όπως το Random Forest, πιθανόν να είναι πιο κατάλληλο για την πρόβλεψη της διάρκειας παραμονής λόγω της πολυπλοκότητας των παραγόντων στο κλινικό περιβάλλον.
- Η μεταβλητή “facid” φαίνεται να έχει σημαντικό ρόλο, καθώς η παρουσία της μπορεί να βοηθήσει στην ακριβέστερη πρόβλεψη με βάση τις διαφορές μεταξύ κλινικών.
- Η απόδοση του μοντέλου χωρίς τη “facid” θα ήταν λιγότερο αξιόπιστη, κάτι που στην πράξη θα μπορούσε να οδηγήσει σε ανακριβείς προβλέψεις και προβλήματα στη διαχείριση πόρων.
-