

**Εργασία στο μάθημα Συστήματα Λήψης
Αποφάσεων
Έτος 2022**

Google Colab URL:

<https://colab.research.google.com/drive/1BD0MwY0pGVm1YzGdjLz610NHCJ8beo14?usp=sharing>

Θέμα:

Ρωσικές απώλειες εξοπλισμού στον πόλεμο Ρωσία-Ουκρανία

Στην εργασία αυτή ασχολήθηκα με τις απώλειες της Ρωσίας όσον αφορά τον εξοπλισμό της την περίοδο 25-02-2022 έως 10-06-2022, στον πόλεμο με την Ουκρανία. Δεν ασχολήθηκα με τα δεδομένα '**personnel**' που περιγράφουν τις απώλειες ανθρώπινου δυναμικού, διότι παρατήρησα ότι έχουν αναλυθεί εκτενώς στο site της kaggle που μας δώσατε. Επίσης, απέρριψα τα δεδομένα '**pow**' (prisoners of war), διότι είχαν από ελάχιστες ως μηδενικές μεταβολές κατά τη διάρκεια του πολέμου, με εξαίρεση την απότομη αύξηση τις πρώτες μέρες. Επομένως, τα δεδομένα που ασχολήθηκα αφορούν μόνο τον εξοπλισμό.

Μέσω της διαδικασίας **Exploratory Analysis** που εφάρμοσα στα δεδομένα, παρατηρούμε μερικές ανωμαλίες σε κάποια από αυτά, οι οποίες δεν τα κάνουν ιδανικά για ανάλυση. Αυτές οι κατηγορίες δεδομένων είναι οι '**vehicles and fuel tanks**' και '**APC**', καθώς βλέπουμε μια απότομη αύξηση στον αριθμό των απωλών εξοπλισμού με μηδενικές προηγούμενες τιμές. Επομένως, κρίνοντας από το HeatMap με τους μέσους όρους κάθε 5 ημερών, οι τρεις κατηγορίες που διάλεξα ως features είναι οι '**tank**', '**drone**' και '**field artillery**'. Αυτό, επειδή φαίνεται να παρουσιάζουν περισσότερη ομαλότητα στις τιμές σχετικά με τις υπόλοιπες κατηγορίες. Θα μπορούσα να διαλέξω και την κατηγορία '**military auto**', ωστόσο τους τελευταίους μήνες παρατηρούμε μηδενικές τιμές.

Τα μοντέλα που επέλεξα να χρησιμοποιήσω είναι τα Theta και Prophet. Ο λόγος αφορά κυρίως στο ότι στην ανάλυση στο site της Kaggle στα δεδομένα 'personnel', τα δύο αυτά μοντέλα φαίνεται να έχουν καλή απόδοση. Ωστόσο, κρίνοντας από το μικρό μέγεθος του dataset με το οποίο ασχολήθηκα, νομίζω ότι τα περισσότερα μοντέλα θα δυσκολευτούν να εξάγουν εύστοχες προβλέψεις.

Τα αποτελέσματα που δώσανε τα δύο μοντέλα για την πρόβλεψη των τελευταίων 10 ημερών του dataset δεν ήταν και πολύ αποδοτικά σε σχέση με τα πραγματικά δεδομένα. Συγκρίνοντας τις μετρικές **MAE**, **RMSE** και **f-score** βλέπουμε ότι και στις δύο περιπτώσεις είναι κοντά η μία με την άλλη. Συνολικά, όμως, φαίνεται πως το μοντέλο Prophet υπερτέρησε στα δεδομένα 'tank', ωστόσο το μοντέλο Theta είχε αρκετά καλύτερη απόδοση στα δεδομένα 'drone' και 'field_artillery', παρότι οι τιμές που προέβλεψε σχηματίζουν ευθεία γραμμή, όπως φαίνεται και στα γραφήματα. Εάν αλλάξουμε τις μέρες πρόβλεψης, αλλάζει και η απόδοση των μοντέλων.

Επειδή η απόδοση των παραπάνω μοντέλων δεν με ικανοποίησε, έψαξα να βρω πιο μοντέλο υπάρχει που ικανοποιεί καλύτερα δεδομένα Time Series, τα οποία δεν παρουσιάζουν ούτε trend ούτε seasonality, όπως στη δική μας περίπτωση. Με λίγο ψάξιμο, έφτασα στο συμπέρασμα ότι η μέθοδος Simple Moving Average (SMA) είναι πιθανών η πιο αποτελεσματική για τέτοιου είδους προβλέψεις, παρότι σχηματίζει ευθεία γραμμή, όπως και το μοντέλο Theta. Πράγματι, υλοποιώντας το μοντέλο φαίνεται πως οι μετρικές MAE και RMSE είναι αισθητά χαμηλότερες από τα προηγούμενα μοντέλα, γεγονός που το καθιστά το πιο accurate.

Συμπερασματικά, πιστεύω ότι αν το dataset ήταν λίγο ως αρκετά μεγαλύτερο, θα μπορούσαμε να εξάγουμε πιο ακριβή αποτελέσματα. Ας ελπίζουμε ότι θα σταματήσει να χρήζει ανανέωσης το δυνατόν συντομότερα.