Intrusion Detection System για IoT Δίκτυα με ML

Ντέιτς Νικόλαος

**Εισαγωγή**

Η αύξηση του αριθμού IoT συσκευών έχει οδηγήσει σε νέες προκλήσεις για την ασφάλεια των δικτύων. Η εργασία αυτή αναπτύσσει ένα IDS βασισμένο σε Machine Learning, εμπνευσμένο από το Saheed et al. (2022), προσαρμοσμένο στο σύγχρονο dataset UNSW-NB15.

**Σύντομη επισκόπηση state-of-the-art**

Το Saheed et al. (2022) πρότεινε IDS με Min-Max normalization, PCA με 10 principal components και έξι ML αλγορίθμους (XGBoost, CatBoost, KNN, SVM, QDA, NB) πετυχαίνοντας accuracy 99.99%. Το UNSW-NB15 χρησιμοποιείται καθώς περιλαμβάνει επιθέσεις που αντικατοπτρίζουν πιο ρεαλιστικά IoT περιβάλλοντα.

**Dataset & Παράμετροι**

Το **UNSW-NB15** είναι ένα σύγχρονο και εκτενές dataset που δημιουργήθηκε για την έρευνα και την ανάπτυξη συστημάτων ανίχνευσης εισβολών (Intrusion Detection Systems – IDS). Δημιουργήθηκε από το Australian Centre for Cyber Security (ACCS) στο Πανεπιστήμιο της Νέας Νότιας Ουαλίας (UNSW) το 2015, με στόχο να ξεπεράσει τους περιορισμούς παλαιότερων συνόλων δεδομένων όπως το KDD99, προσφέροντας πιο ρεαλιστικά σενάρια επιθέσεων και κανονικής κίνησης.

Η διαδικασία δημιουργίας του dataset περιλάμβανε την παραγωγή κίνησης δικτύου μέσω του εργαλείου IXIA PerfectStorm. Το αποτέλεσμα ήταν ένα μείγμα πραγματικής και συνθετικής κυκλοφορίας, που καλύπτει μεγάλο εύρος επιθέσεων και φυσιολογικής δραστηριότητας. Στη συνέχεια, τα πακέτα καταγράφηκαν σε μορφή pcap και αναλύθηκαν με το Bro-IDS (Zeek) και το Argus για την εξαγωγή χαρακτηριστικών σε επίπεδο ροής.

Το τελικό dataset περιέχει **257,673 εγγραφές**, οι οποίες χωρίζονται σε training και testing υποσύνολα. Κάθε εγγραφή περιγράφει μια σύνδεση δικτύου με **49 χαρακτηριστικά (features)**, όπως:

* βασικές πληροφορίες σύνδεσης (π.χ. διάρκεια, πρωτόκολλο),
* στατιστικά στοιχεία (π.χ. αριθμός bytes και packets),
* χρονικές μετρήσεις και συμπεριφορικά χαρακτηριστικά.

Κάθε εγγραφή φέρει μία ετικέτα που δηλώνει αν πρόκειται για κανονική δραστηριότητα ή για κακόβουλη ενέργεια. Οι επιθέσεις ταξινομούνται σε **εννέα κατηγορίες**, μεταξύ των οποίων περιλαμβάνονται:

* Fuzzers,
* DoS (Denial of Service),
* Exploits,
* Generic,
* Reconnaissance,
* Analysis,
* Backdoor,
* Shellcode,
* Worms.

Η πλούσια θεματολογία και η ποικιλία των επιθέσεων καθιστούν το UNSW-NB15 ιδανικό για εκπαίδευση και αξιολόγηση αλγορίθμων μηχανικής μάθησης, καθώς και για μελέτη τεχνικών ανίχνευσης δικτυακών απειλών.

Συνοπτικά, το UNSW-NB15 θεωρείται σήμερα ένα από τα πιο αξιόπιστα benchmark datasets για ερευνητικές εργασίες στον τομέα της κυβερνοασφάλειας και έχει χρησιμοποιηθεί εκτενώς σε πειράματα που αφορούν:

* ανίχνευση ανωμαλιών (anomaly detection),
* ταξινόμηση επιθέσεων,
* μεθόδους μείωσης διαστάσεων,
* συστήματα real-time ανίχνευσης.

Το dataset έγινε reshuffle και χωρίστηκε εκ νεου σε training (80%) και testing (20%)(validation – testing για το νευρωνικό) καθώς στο dataset του Kaggle το training split είναι μικρότερο από το test split. Χαρακτηριστικά από IP/TCP headers και payload χρησιμοποιήθηκαν για ανίχνευση επιθέσεων.

Προεπεξεργασία & Επιλογή χαρακτηριστικών

Εφαρμόστηκε MinMaxScaler για κανονικοποίηση τιμών και one hot encoding των μη αριθμητικών μεταβλητών.Όλα τα μοντέλα (εκτός του νευρωνικού το οποίο έτρεξε μόνο σε ολόκληρο το σετ δεδομένων) δοκιμάστηκαν σε 3 διαφορετικές εκδοχές του dataset.

* Full dataset
* Σε επιλεγμένα features του dataset.Συγκεκριμένα χρησιμοποιήθηκε Random Forest για επιλογή χαρακτηριστικών με σημαντικότητα πάνω από τη διάμεσο.
* PCA με 30 principal components.

**Μοντέλα & Υλοποίηση**

Εκπαιδεύτηκαν έξι μοντέλα: FNN, AdaBoost, XGBoost, Logistic Regression, Random Forest και stacked model που συνδύασε τα παραπάνω με meta-classifier Logistic Regression. Η υλοποίηση έγινε σε Python με scikit-learn, XGBoost και Pytorch.

Συγκεκριμένα τα αποτελέσματα που παρουσιάζονται αφορούν τα εξής μοντέλα :

* FNN με 6 fc layers trained για 20 epochs με RELU activation, dropout=0.2 στα layers 4 και 5 and Batchnorm.
* AdaBoost με maxdepth=3 , 200 estimators και 0.1 learning rate.
* XGboost με maxdepth=6 , 300 estimators και 0.2 learning rate.
* Logistic Regression με L1 Regularizatiion ,C 0.01 and 1000 max iterations.
* Random Forest Classifier με 100 estimators ,min\_samples\_split 2 και min samples leaf 1
* Stacked model από Adaboost,XGBoost,Random Forest and LR με Logistic Regression on top of voting ως metaclassifier.

# Πειράματα & Αποτελέσματα

Τα αποτελέσματα τεκμηριώνουν ότι το μοντέλο Stacked\_selected εμφανίζει οριακά την υψηλότερη συνολική απόδοση, επιτυγχάνοντας ακρίβεια 0.95097, ενώ οι υπόλοιπες μετρικές κυμαίνονται σε παρόμοιο επίπεδο (precision: ~0.951, recall: ~0.951, F1 score: ~0.951), καταδεικνύοντας ισχυρή ισορροπία μεταξύ θετικής πρόβλεψης και ανάκτησης.

Το xgb\_selected καταγράφει σχεδόν ταυτόσημη συμπεριφορά, με ελαφρώς υψηλότερο precision (0.95124) και F1 score (0.95103), γεγονός που υποδεικνύει υψηλή συνέπεια και αξιοπιστία στην ταξινομητική διαδικασία. Στην περίπτωση των αντίστοιχων μη βελτιστοποιημένων εκδόσεων (xgb και Stacked), παρατηρείται μέτρια πτώση στην ακρίβεια (0.95044 και 0.94881 αντίστοιχα) και αντίστοιχα οριακές μειώσεις σε precision και recall. Η διαφορά αυτή επιβεβαιώνει τη συνεισφορά της feature selection στην αύξηση των ταξινομητικών επιδόσεων.

Το FNN υπολείπεται σε όλα τα κριτήρια (accuracy: 0.93441, F1 score: 0.93363), φανερώνοντας χαμηλότερη ικανότητα γενίκευσης, ενδεχομένως λόγω περιορισμένης χωρητικότητας του μοντέλου ή ανεπαρκούς συντονισμού υπερπαραμέτρων.

Τα μοντέλα ada και ada\_selected παρουσιάζουν σχεδόν ταυτόσημα αποτελέσματα (accuracy: 0.91571, F1 score: 0.91492), στοιχείο που υποδηλώνει ότι η επιλογή χαρακτηριστικών σε αυτό το πλαίσιο δεν προσφέρει ουσιαστική βελτίωση. Αντίθετα, η ενσωμάτωση PCA στα αντίστοιχα μοντέλα (Stacked\_PCA, xgb\_PCA, ada\_PCA) συνοδεύεται από αξιοσημείωτη μείωση όλων των μετρικών, με το ada\_PCA να καταγράφει τη χαμηλότερη συνολική επίδοση (accuracy: 0.85726, precision: 0.85231, recall: 0.84769, F1 score: 0.84965). Το αποτέλεσμα αυτό υποδηλώνει ότι η μείωση διαστάσεων με PCA πιθανώς απομάκρυνε πληροφορία υψηλής διακριτικής αξίας για το πρόβλημα ταξινόμησης.

Συνολικά οι τεχνικές stacking και feature selection προσδίδουν βελτιωμένα αποτελέσματα ως προς την ακρίβεια και την ισορροπία precision–recall–F1. Αντίθετα, η χρήση PCA μειώνει συστηματικά την αποτελεσματικότητα των μοντέλων, στοιχείο που πρέπει να ληφθεί υπόψη σε περιπτώσεις με υψηλή διαστασιακότητα, όπου το trade-off μεταξύ μείωσης πολυπλοκότητας και διατήρησης πληροφορίας είναι κρίσιμο.

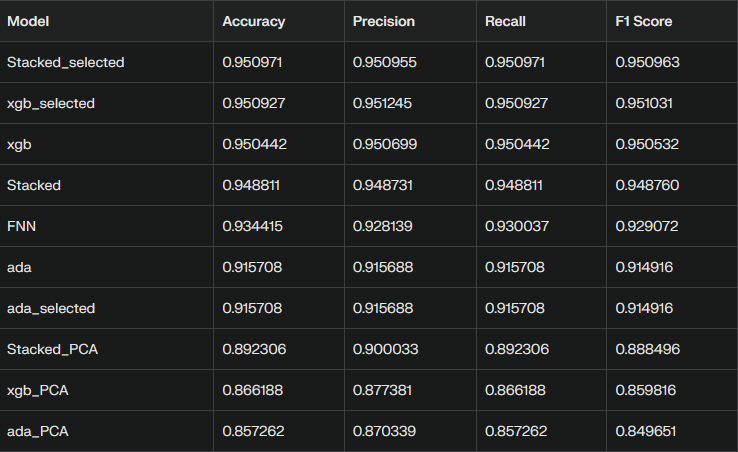
**Πίνακες αποτελεσμάτων**

Τα αποτελέσματα του paper:

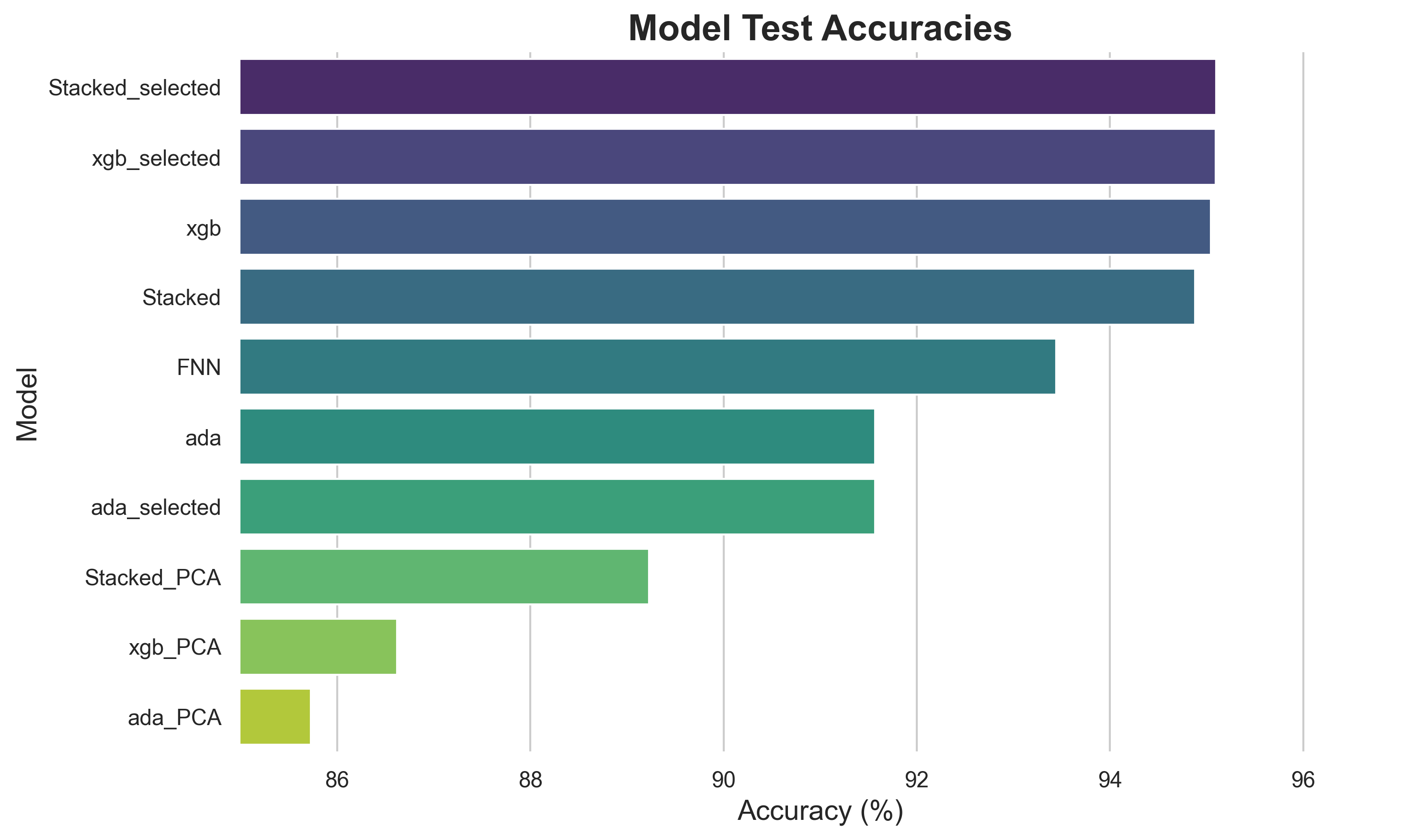
A screenshot of a computer screen

AI-generated content may be incorrect.

Τα αποτελέσματα της δικής μας ανάλυσης:



Γραφική αναπαράσταση αποτελεσμάτων:



**Συγκριτική αξιολόγηση**

Η εργασία προσφέρει διαφορετική προσέγγιση, χρησιμοποιώντας συνδυασμό μοντέλων και επιλογή χαρακτηριστικών με Random Forest. Παρόλο που το accuracy ήταν χαμηλότερο (95% vs 99.99%), η ανάλυση καλύπτει περισσότερες παραλλαγές και αναδεικνύει τη σημασία επιλογής χαρακτηριστικών. Παρότι στην αρχή δοκιμάστηκαν τα documented βήματα που ακολουθήθηκαν στο paper τα αποτελέσματα διέφεραν κατά πολύ. Καταλαβαίνουμε οτι υπάρχουν βήματα πιθανώς στο preprocessing τα οποία δεν είναι documented.