

## Σύγκριση Α και Β μέρους

### Αποτελέσματα

| Αλγόριθμος                      | Precision<br>(macro) | Recall (macro) | F1 (macro) | Accuracy |
|---------------------------------|----------------------|----------------|------------|----------|
| Naive Bayes                     | 0.843                | 0.841          | 0.841      | 84.1%    |
| Logistic<br>Regression<br>(SGD) | 0.865                | 0.865          | 0.865      | 86.5%    |
| AdaBoost                        | 0.774                | 0.770          | 0.770      | 77.0%    |
| BiGRU + Global<br>Max Pooling   | 0.860                | 0.856          | 0.855      | 85.6%    |

### Σχολιασμός αποτελεσμάτων

Με βάση τα παραπάνω, η καλύτερη συνολική επίδοση εμφανίζεται στη Logistic Regression, η οποία πέτυχε accuracy 86.5%. Το συγκεκριμένο μοντέλο χρησιμοποιεί γραμμικό ταξινομητή και βασίζεται σε binary bag-of-words αναπαράσταση, μαζί με επιλογή χαρακτηριστικών (Information Gain), κάτι που φαίνεται να λειτουργεί πολύ καλά για το IMDB.

Το BiGRU ήρθε πολύ κοντά (85.6%), αλλά δεν κατάφερε να ξεπεράσει την Logistic Regression. Παρόλα αυτά, είναι αρκετά ανταγωνιστικό και έχει το πλεονέκτημα ότι εκμεταλλεύεται σημασιολογικές αναπαραστάσεις μέσω pre-trained GloVe embeddings και κρατάει τη σειρά των λέξεων.

Ο Naive Bayes εμφανίζει επίσης καλή απόδοση (84.1%), κάτι που είναι αναμενόμενο μιας και πρόκειται για ένα απλό αλλά αποτελεσματικό baseline. Τέλος, το AdaBoost είχε αισθητά χαμηλότερα αποτελέσματα (77.0%), πιθανότατα επειδή χρειάζεται πιο προσεκτικό tuning ή διαφορετική επιλογή χαρακτηριστικών.

### Σύγκριση προσεγγίσεων

Στο Μέρος Α χρησιμοποιήθηκε μια πιο “παραδοσιακή” προσέγγιση με binary bag-of-words αναπαράσταση. Πριν την εκπαίδευση, έγινε preprocessing με αφαίρεση πολύ συχνών και πολύ σπάνιων λέξεων, ενώ η τελική επιλογή περιλάμβανε τις 5000 πιο σημαντικές λέξεις σύμφωνα με το Information Gain. Το βασικό πλεονέκτημα αυτής της μεθόδου είναι ότι είναι γρήγορη και σχετικά εύκολη στην υλοποίηση και ερμηνεία. Ωστόσο, χάνει τη σειρά των λέξεων και δεν “καταλαβαίνει” τη σημασία τους.

Στο Μέρος B, το μοντέλο BiGRU αξιοποιεί embeddings από το GloVe (100 διαστάσεων) και αποτελείται από δύο bidirectional GRU layers. Η χρήση Global Max Pooling βοηθάει στη συμπύκνωση της πληροφορίας από όλη την πρόταση. Παρότι είναι πιο σύνθετο και αργό στη εκπαίδευση (συνήθως χρειάζεται GPU για να τρέχει άνετα), έχει το πλεονέκτημα ότι μπορεί να μάθει πιο πλούσιες σημασιολογικές αναπαραστάσεις και να αξιοποιεί τη σειρά των λέξεων.

## Συμπεράσματα

- Η Logistic Regression πέτυχε την καλύτερη επίδοση (86.5%), ξεπερνώντας οριακά το BiGRU.
- Το BiGRU είναι πολύ κοντά σε απόδοση, αλλά δεν παρουσιάζει ξεκάθαρη υπεροχή στο συγκεκριμένο dataset.
- Ένας λόγος που τα deep μοντέλα δεν “ξεφεύγουν” εδώ είναι ότι το IMDB έχει σχετικά περιορισμένο μέγεθος, ενώ το bag-of-words με σωστό feature selection αποτελεί ήδη πολύ δυνατό baseline.
- Σε επίπεδο πρακτικής χρήσης, η Logistic Regression είναι πιο γρήγορη και απλή. Από την άλλη, το BiGRU μπορεί να αποδώσει καλύτερα όταν υπάρχουν περισσότερα δεδομένα ή γίνεται πιο επιθετικό tuning/fine-tuning.

## Υπερπαράμετροι BiGRU

| Παράμετρος          | Τιμή                  |
|---------------------|-----------------------|
| Embedding Dimension | 100 (GloVe)           |
| Hidden Dimension    | 128                   |
| GRU Layers          | 2                     |
| Bidirectional       | True                  |
| Dropout             | 0.5                   |
| Optimizer           | Adam                  |
| Learning Rate       | 0.001                 |
| Batch Size          | 64                    |
| Best Epoch          | 5 (based on dev loss) |