

--ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ, ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΡΟΜΠΟΤΙΚΗ»

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
ΠΕΤΣΙΟΣ ΝΙΚΟΛΑΟΣ
ΑΜ:55

«ΑΝΑΓΝΩΡΙΣΗ ΤΟΥ ΕΙΔΟΥΣ ΜΙΑΣ ΤΑΙΝΙΑΣ ΑΠΟ
TRAILER»

ΕΠΙΒΛΕΠΩΝ :ΝΙΚΟΛΑΙΔΗΣ ΑΘΑΝΑΣΙΟΣ

ΣΕΠΤΕΜΒΡΙΟΣ 2021

<ΕΥΧΑΡΙΣΤΙΕΣ>

Η συγκεκριμένη διπλωματική εργασία πραγματοποιήθηκε στα πλαίσια του μεταπτυχιακού κύκλου σπουδών του τμήματος Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών του Διεθνές Πανεπιστημίου της Ελλάδος κι έχει ως θέμα: «Αναγνώριση του είδους μιας ταινίας από trailer».

Πριν την παρουσίαση των αποτελεσμάτων της παρούσας διπλωματικής εργασίας, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή

κ. Αθανάσιο Νικολαΐδη

για την πολύτιμη βοήθεια, εμπιστοσύνη και καθοδήγησή τους καθ' όλη τη διάρκεια της δουλειάς μου, ο οποίος έπαιξε καθοριστικό ρόλο στην πραγματοποίησή της. Τον ευχαριστώ για την προθυμία, τις συμβουλές, τις υποδείξεις, τις διορθώσεις, τις επισημάνσεις και για τις γνώσεις που μου παρείχε. Τον ευχαριστώ επίσης διότι μου έδωσε τη δυνατότητα να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα, χρησιμοποιώντας την γλώσσα Python, το οποίο με την ολοκλήρωσή του αποτελεί μεγάλη προσωπική ικανοποίηση.

Τέλος αισθάνομαι την ανάγκη να ευχαριστήσω την οικογένεια μου για την ηθική συμπαράσταση που μου έδειξαν κατά τη διάρκεια ολοκλήρωσης της διπλωματικής μου εργασίας.

Περιεχόμενα

| | |
|--|----|
| 1. Εισαγωγή | 10 |
| 1.1 Αντικείμενο Εργασίας | 12 |
| 1.2 Δομή Εργασίας | 14 |
| 2. Είδη Ταινιών | 15 |
| 2.1 Περιγραφή Ειδών Ταινιών | 19 |
| 2.2 Αυτόματη Αναγνώριση Ειδών Ταινιών | 19 |
| 2.2.1 Βιβλιογραφική Ανασκόπηση | 19 |
| 3. Μεθοδολογία Αναγνώρισης Είδους Ταινίας από Trailer | 23 |
| 3.1 Σύνολο Δεδομένων | 26 |
| 3.1.1 Περιγραφή Συνόλου Δεδομένων | 26 |
| 3.1.2 Προεπεξεργασία Δεδομένων | 29 |
| 3.1.3 Ανίχνευση Αντικειμένων | 30 |
| 3.1.3.1. YOLO | 31 |
| 3.1.4 Διαμόρφωση Συνόλου Δεδομένων Ταξινόμησης | 32 |
| 3.2 Ταξινόμηση του Είδους των Ταινιών | 33 |
| 3.2.1 Multi-label Ταξινόμηση | 34 |
| 3.2.2 Προεπεξεργασία | 35 |
| 3.2.3 Αλγόριθμοι | 39 |
| 4. Αποτελέσματα Αξιολόγησης Αλγορίθμων | 45 |
| 4.1 Μετρικές Αξιολόγησης | 45 |
| 4.1.1 Binary Ταξινόμηση | 45 |
| 4.1.2 Multi-label Ταξινόμηση | 47 |
| 4.2 Αποτελέσματα | 48 |
| 4.2.1 Αξιολόγηση στο Multi-Label Πρόβλημα Ταξινόμησης | 49 |
| 4.2.2 Αξιολόγηση στα Επιμέρους Binary Προβλήματα Ταξινόμησης | 52 |
| 4.2.3 Εκπαίδευση και Παραμετροποίηση του Νευρωνικού Δικτύου | 60 |
| 4.2.4 Επιλογή Σημαντικότερων Χαρακτηριστικών | 61 |
| 4.3 Σύγκριση Αποτελεσμάτων | 67 |
| 5. Συμπεράσματα | 69 |
| Βιβλιογραφία | 72 |

Κατάλογος Σχημάτων

| | |
|--|----|
| Σχήμα 1.1: Οι δαπάνες σε δισεκατομμύρια δολάρια για περιεχόμενο και οι συνδρομητές επί πληρωμή του Netflix σε εκατομμύρια (Joo, 2020)..... | 11 |
| Σχήμα 3.1: Η μεθοδολογία που έχει ακολουθηθεί για την αυτόματη αναγνώριση του είδους μιας ταινίας..... | 25 |
| Σχήμα 3.2: Η κατανομή της διάρκειας και των σκηνών στα trailers του MovieNet (Huang et al., n.d.). | 27 |
| Σχήμα 3.3: Το πλήθος των ταινιών ανά κατηγορία στο αρχικό σύνολο δεδομένων που συλλέγεται..... | 28 |
| Σχήμα 3.4: Το πλήθος των ταινιών συναρτήσει του πλήθους των κατηγοριών στο οποίο ανήκουν στο αρχικό σύνολο δεδομένων που συλλέγεται..... | 29 |
| Σχήμα 3.5: Ιστόγραμμα του πλήθους των frames στις ταινίες που απαρτίζουν το αρχικό σύνολο δεδομένων..... | 29 |
| Σχήμα 3.6: Το πλήθος των ταινιών συναρτήσει του πλήθους των κατηγοριών στο οποίο ανήκουν τελικό σύνολο δεδομένων προς ταξινόμηση..... | 38 |
| Σχήμα 3.7: Ιστόγραμμα του πλήθους των frames στις ταινίες που απαρτίζουν το τελικό σύνολο δεδομένων προς ταξινόμηση..... | 38 |
| Σχήμα 3.8: Μία ενδεικτική αρχιτεκτονική ενός νευρωνικού δικτύου..... | 44 |
| Σχήμα 4.1: Η δομή ενός confusion matrix, που χρησιμοποιείται για το χαρακτηρισμό των προβλέψεων ενός αλγορίθμου ταξινόμησης..... | 46 |
| Σχήμα 4.2: Η κατανομή των δειγμάτων στις 5 διαφορετικές κλάσεις στο training set..... | 49 |
| Σχήμα 4.3: Η κατανομή των δειγμάτων στις 5 διαφορετικές κλάσεις στο test set..... | 49 |
| Σχήμα 4.4: Η καμπύλη Precision-Recall που προκύπτει από τον αλγόριθμο SVM για κάθε κλάση..... | 51 |
| Σχήμα 4.5: Τα confusion matrices που προκύπτουν από την αξιολόγηση του Random Forest για κάθε είδος ταινίας..... | 57 |
| Σχήμα 4.6: Η απόδοση των αλγορίθμων για τις διαφορετικές αρχιτεκτονικές MLP. Με τις πλήρεις γραμμές εμφανίζεται το accuracy κατά την εκπαίδευση, ενώ με τις διακεκομμένες το accuracy κατά την αξιολόγηση..... | 58 |
| Σχήμα 4.7: Η συχνότητα εμφάνισης των χαρακτηριστικών που προκύπτουν τουλάχιστον μία φορά ως ένα από τα 10 σημαντικότερα για κάθε είδος ταινίας βάσει της δοκιμασίας χ^2 | 60 |
| Σχήμα 4.8: Ραβδόγραμμα που εμφανίζει τη σημασία του συνόλου των χαρακτηριστικών που προκύπτει ως η ένωση των 10 σημαντικότερων χαρακτηριστικών ανά είδος ταινίας, βάσει του Random Forest..... | 61 |
| Σχήμα 4.9: Το δέντρο αποφάσεων για το είδος Drama..... | 63 |
| Σχήμα 4.10: Το δέντρο αποφάσεων για το είδος Comedy..... | 64 |
| Σχήμα 4.11: Το δέντρο αποφάσεων για το είδος Romance..... | 64 |
| Σχήμα 4.12: Το δέντρο αποφάσεων για το είδος Action..... | 65 |
| Σχήμα 4.13: Το δέντρο αποφάσεων για το είδος Thriller..... | 65 |

Κατάλογος Πινάκων

| | |
|---|----|
| Πίνακας 2.1: Το προτιμότερο είδος ταινίας ανά χώρα (New York Film Academy, 2016)..... | 18 |
| Πίνακας 3.1: Ένα δείγμα του συνόλου δεδομένων, όπως διαμορφώνεται, ώστε να ακολουθήσει η ταξινόμηση των ειδών μίας ταινίας..... | 32 |
| Πίνακας 3.2: Κατηγορίες ταινιών στο αρχικό σύνολο δεδομένων που συλλέγεται, προτού υποστή επεξεργασία..... | 37 |
| Πίνακας 3.3: Ένα δείγμα του τελικού συνόλου δεδομένων, όπως διαμορφώνεται, ώστε να ακολουθήσει η ταξινόμηση των ειδών μίας ταινίας..... | 39 |
| Πίνακας 3.4: Οι αλγόριθμοι ταξινόμησης και οι αντίστοιχες παράμετροι που χρησιμοποιούνται κατά τον πειραματισμό με το σύνολο δεδομένων..... | 44 |
| Πίνακας 4.1: Η συνολική απόδοση των αλγορίθμων..... | 50 |
| Πίνακας 4.2: Οι παράμετροι των αλγορίθμων ταξινόμησης που επιτυγχάνουν τα καλύτερα αποτελέσματα..... | 51 |
| Πίνακας 4.3: Η απόδοση των αλγορίθμων για τις ταινίες Drama..... | 52 |
| Πίνακας 4.4: Η απόδοση των αλγορίθμων για τις ταινίες Comedy..... | 53 |
| Πίνακας 4.5: Η απόδοση των αλγορίθμων για τις ταινίες Romance..... | 54 |
| Πίνακας 4.6: Η απόδοση των αλγορίθμων για τις ταινίες Action..... | 55 |
| Πίνακας 4.7: Η απόδοση των αλγορίθμων για τις ταινίες Thriller..... | 56 |
| Πίνακας 4.8: Τα 10 σημαντικότερα χαρακτηριστικά ανά είδος ταινίας, σύμφωνα με τη δοκιμασία χ^2 στο training set..... | 59 |

Κατάλογος Συντομογραφιών

| | |
|------|------------------------------|
| CNN | Convolutional Neural Network |
| COCO | Common Objects in Context |
| CSV | Comma-Separated Values |
| DNN | Deep Neural Network |
| MLP | Multi-layer Perceptron |
| RNN | Recurrent Neural Networks |
| SVM | Support Vector Machine |
| YOLO | You Only Look Once |
| TN | Τεχνητή Νοημοσύνη |

Περίληψη

Η αυτοματοποίηση πλήθους διαδικασιών με τη χρήση τεχνικών Τεχνητής Νοημοσύνης κερδίζει ολοένα και περισσότερο έδαφος. Μια διαδικασία η οποία φαίνεται να κεντρίζει το ενδιαφέρον των ερευνητών είναι η αυτόματη αναγνώριση του είδους μιας ταινίας. Λαμβάνοντας υπόψη πως η παραγωγή ταινιών αποτελεί μία από τις μεγαλύτερες βιομηχανίες παγκοσμίως και πως η παρακολούθηση αυτών προτιμάται ως ψυχαγωγική δραστηριότητα από το μεγαλύτερο μέρος του πληθυσμού για αξιοποίηση του ελεύθερου χρόνου, μπορεί να αιτιολογηθεί αυτή η τάση για τη συγκεκριμένη αυτοματοποίηση.

Σε γενικά πλαίσια, το είδος μιας ταινίας αποτελεί τον κύριο χαρακτηρισμό της και το πιο αντιπροσωπευτικό στοιχείο για την κατηγοριοποίηση της σε σύγκριση με άλλες. Το είδος είναι αυτό το οποίο μπορεί να προσελκύσει το κοινό που επιδεικνύει μια προτίμηση σε συγκεκριμένα σενάρια, σκηνικά και διαλόγους που αντιστοιχούν στο ανάλογο είδος. Επιπλέον, ανάμεσα στο πλήθος των ποικίλων ειδών υπάρχουν κάποια που χαίρουν της προτίμησης της πλειονότητας των θεατών, προσδιορίζοντας, κατ' επέκτασιν, το επίκεντρο των κινηματογραφικών παραγωγών.

Η αυτόματη αναγνώριση του είδους μιας ταινίας μπορεί να στηριχθεί σε δεδομένα που προέρχονται κυρίως από προωθητικές ενέργειες αυτής όπως η αφίσα της, ένα clip και κυρίως το trailer της. Στην εν λόγω εργασία χρησιμοποιούνται τα trailers, η συλλογή των οποίων μέσω του IMDb και του YouTube συντελεί στην κατασκευή ενός συνόλου δεδομένων με 209 trailers. Στα πλαίσια της προεπεξεργασίας του συνόλου αυτού, εφαρμόζεται μία τεχνική ανίχνευσης αντικειμένων που ονομάζεται YOLO, η οποία αναγνωρίζει το είδος των αντικειμένων και τη συχνότητα εμφάνισής τους σε σκηνές του trailer της εκάστοτε ταινίας.

Τα αποτελέσματα της προεπεξεργασίας αυτής δίνουν ένα ακόμη σύνολο δεδομένων, το οποίο χρησιμοποιείται ως είσοδος για την εκπαίδευση ευφυών αλγορίθμων Τεχνητής Νοημοσύνης με σκοπό την ταξινόμηση της ταινίας βάσει είδους. Ενδιαφέρον παρουσιάζει το γεγονός ότι το συγκεκριμένο πρόβλημα ταξινόμησης χαρακτηρίζεται ως multi-label, με δεδομένο ότι μια ταινία μπορεί να ανήκει σε περισσότερα από ένα είδη.

Με στόχο την επίτευξη υψηλότερων αποδόσεων, πραγματοποιείται εκτενής πειραματισμός με διαφορετικούς αλγορίθμους και παραμέτρους. Τα αποτελέσματα φανερώνουν την καταλληλότητα της μεθοδολογίας που χρησιμοποιείται στη παρούσα εργασία, επιτυγχάνοντας αρκετά καλές επιδόσεις. Παράλληλα, παρέχεται μια ερμηνεία και επεξήγηση για τα χαρακτηριστικά (επί της ουσίας τα αντικείμενα) που διαδραματίζουν το σημαντικότερο ρόλο στην αναγνώριση του είδους μιας ταινίας. Τέλος, η μεθοδολογία και τα αποτελέσματά της συγκρίνονται με τη βιβλιογραφία.

Abstract

The automation of a number of processes using Artificial Intelligence techniques is gaining more and more ground. One process that seems to be of interest to researchers is the automatic recognition of the genre of a film. Given that film production is one of the largest industries in the world and that watching it is preferred as a leisure activity by the majority of the population, the tendency for this automation can be justified.

In general, the genre of a film is its main characterization and the most representative element for its categorization compared to others. The genre is the one that can attract the audience that shows a preference for specific scenarios, scenes and dialogues that correspond to the specific genre. In addition, among the multitude of different genres, there are some that enjoy the preference of the majority of viewers, thus determining the focus of film producers.

The automatic recognition of the genre of a movie can be based on data that comes mainly from its promotion, such as its poster, a clip and especially its trailer. In this work, trailers are employed, the collection of which through IMDb and YouTube contributes to the construction of a dataset with 209 trailers. As part of the pre-processing of this dataset, an object detection technique called YOLO is applied, which determines the type of objects and the frequency of their appearance in scenes of the trailer of each film.

The results of this pre-processing provide another dataset, which is used as input for the training of intelligent Artificial Intelligence algorithms and in general the classification of the type of film. It is interesting that this classification problem is characterized as multi-label, given that a film can belong to more than one genre.

In order to achieve higher evaluation results, extensive experimentation is carried out with different algorithms and parameters. The results show the appropriateness of the methodology used in the present work, achieving an acceptable performance. At the same time, an interpretation and explanation is provided for the characteristics (in essence objects) that play the most important role in the recognition of the genre of a film. Last but not least, the methodology and acquired results are compared with the literature.

1. Εισαγωγή

Η παρακολούθηση ταινιών αποτελεί ένα σχεδόν αναπόσπαστο κομμάτι της καθημερινότητας των ανθρώπων. Μέσω μια ταινίας, ο άνθρωπος μπορεί να χαλαρώσει, να αποκτήσει γνώσεις, αλλά και να περάσει ευχάριστα και ορισμένες φορές εποικοδομητικά τον ελεύθερο χρόνο του. Για κάθε έναν από τους προαναφερόμενους σκοπούς, υπάρχει και ένα είδος ταινίας που τον ικανοποιεί, γεγονός που θα αναπτυχθεί κι εξηγηθεί περαιτέρω σε επόμενο κεφάλαιο.

Οι ταινίες, ακόμα και αυτές που σχετίζονται με τη βιογραφία ενός αξιόλογου ατόμου, περιγράφουν όχι απόλυτα ρεαλιστικές έως και πλήρως φανταστικές καταστάσεις, κάνοντας πολλές φορές χρήση μη-υπαρκτών προσώπων. Κύριο μέλημά τους αποτελεί η ψυχαγωγία του θεατή, η ταύτισή του με τους πρωταγωνιστικούς ρόλους και η δημιουργία τροφής για σκέψη, σχετικά με διαφορετικές και ποικίλες αντιλήψεις.

Επιπλέον, η παρακολούθηση ταινιών μπορεί να αξιοποιηθεί ως ένα μέσο κοινωνικοποίησης του θεατή και διαμόρφωσης δυνατών συναισθηματικών δεσμών, όταν αυτή πραγματοποιείται με παρέα. Σε αυτήν την περίπτωση, δημιουργείται ένα ακόμα πιο ευχάριστο κλίμα, το οποίο πυροδοτεί συζητήσεις, εξασκώντας παράλληλα την κριτική σκέψη και το νου.

Ωστόσο, είναι σημαντικό να αναφερθεί το γεγονός πως μια ταινία μπορεί να επηρεάσει τη συμπεριφορά ενός ατόμου, γι' αυτό και απαιτείται ιδιαίτερη προσοχή στις σκηνές και στα μηνύματα που περνά. Για παράδειγμα, σκηνές βίας, παρά το γεγονός ότι αποτελούν ένα φανταστικό κατασκεύασμα, μπορούν να οδηγήσουν σε έξαρση βίαιων συμπεριφορών. Βέβαια, κάτι τέτοιο συνήθως συναντάται σε περιπτώσεις, όπου ο θεατής αντιμετωπίζει κάποιες ψυχικές διαταραχές ή είναι μικρός ηλικιακά και μιμείται συμπεριφορές, μη έχοντας την κρίση να κατανοήσει το αντίτιμο των πράξεών του (Nabi & Clark, 2008).

Ομολογουμένως η βιομηχανία του κινηματογράφου είναι πολύ μεγάλη και ισχυρή, γεγονός που αιτιολογεί τον χαρακτηρισμό του κινηματογράφου ως η “έβδομη τέχνη”. Ιδιαίτερα, μετά την επιτυχία των χολιγουντιανών παραγωγών, η ανάπτυξη μιας οικονομίας με βελτίωση του κινηματογράφου αποτελεί πλέον κύριο πυλώνα (Κοντονή, 2015).

Ως βιομηχανία εξελίσσεται, αναπτύσσεται και τροποποιείται με γοργούς ρυθμούς, τόσο ως προς το περιεχόμενο μιας ταινίας και τη διαδικασία παραγωγής της, όσο και στον τρόπο που το κοινό προτιμά να παρακολουθήσει την ταινία της επιλογής του. Ιδιαίτερα τα τελευταία χρόνια εκτός από τους κινηματογράφους παρατηρείται μία άνθιση διάφορων πλατφορμών, συνδρομητικών υπηρεσιών και ιστοσελίδων για την παρακολούθηση διάφορων κινηματογραφικών παραγωγών, χρονολογικά νεότερων αλλά και παλαιότερων.

Το γεγονός αυτό επιβεβαιώνεται αν αναλογιστεί κανείς την εκθετική αύξηση των συνδρομητών της πλατφόρμας Netflix τα τελευταία χρόνια. Ειδικά στην περίπτωση του Netflix, αξίζει να σημειωθεί πως έχει γνωρίσει μεγάλη απήχηση και επιτυχία, παρά το ότι δεν διαθέτει τις παραγωγές της σε κινηματογράφους για προβολή, κάτι που αποτελεί αντικείμενο

μελέτης για πολλές έρευνες (Joo, 2020). Ως προς την αύξηση των συνδρομητών και των εσόδων του Netflix, αυτά απεικονίζονται στο Σχήμα 1.1.

Υπάρχουν πολλά χαρακτηριστικά, τα οποία διέπουν μια ταινία και διαμορφώνουν μια άποψη στον εκάστοτε θεατή, ώστε να τα συμπεριλάβει ως κριτήρια για παρακολούθηση ή αξιολόγηση αυτής. Μεταξύ αυτών, είναι οι ηθοποιοί που κατέχουν τους πρωταγωνιστικούς ρόλους, η εταιρεία, ο τόπος και οι χρηματικές δαπάνες της παραγωγής, ο σκηνοθέτης και πιο συχνά το είδος της ταινίας.

Είθισται η κατηγοριοποίηση των ταινιών σε αντίστοιχα είδη να πραγματοποιείται με γνώμονα την ιδιοσυγκρασία των χαρακτήρων και την ιστορία που περιγράφεται σε αυτές. Τόσο στον κινηματογράφο, όσο και στην τηλεόραση, το είδος μιας κινηματογραφικής ταινίας ή μιας τηλεοπτικής σειράς, αντίστοιχα, χρησιμοποιείται για να εντείνει τις προσδοκίες του κοινού.

Παρά ταύτα, κύριο ερώτημα των επαγγελματιών του κλάδου είναι ποιο από τα προαναφερόμενα χαρακτηριστικά προσελκύει μεγαλύτερη μερίδα θεατών και καθορίζει και μεγαλύτερο ποσοστό επιτυχίας, άρα και την αύξηση εσόδων. Ετησίως, κάνουν την εμφάνισή τους διάφορες νέες θεωρίες αναφορικά με τις προτιμήσεις των θεατών, αν λόγου χάριν ενδιαφέρονται για ταινίες που σχετίζονται με κόμικς, βρικόλακες, υπερφυσικά φαινόμενα, κ.ο.κ.

Κύριο ρόλο στην απήχηση μιας ταινίας διαδραματίζει το trailer αυτής, το οποίο ουσιαστικά προβάλλει στιγμιότυπα και γενικότερα στοιχεία της, μέσω των οποίων μπορεί να προωθηθεί αποτελεσματικότερα η ταινία και να προσελκύσει μεγαλύτερη μερίδα κόσμου. Ένα trailer στοχεύει στη δημιουργία μιας πιο καλλιτεχνικής ατμόσφαιρας, περιλαμβάνοντας σκηνές διαλόγου και αποσπάσματα που διεγείρουν το ενδιαφέρον (Singhal et al., 2018).

Επιπροσθέτως, οι εταιρείες παραγωγής επικεντρώνονται και στον εντοπισμό μερικών ελκυστικών λέξεων-κλειδιών, προκειμένου να συμπεριληφθούν στο αντίστοιχο trailer και να παροτρύνουν το κοινό να προτιμήσει την ταινία και να αγοράσει ένα εισιτήριο (Singh et al., 2019).

| Year | Cash Spending on Content | Total Paid Subscribers (U.S. Subscribers) |
|------|--------------------------|---|
| 2011 | - | 21.6 |
| 2013 | - | 41.4 |
| 2015 | \$4.61 | 70.8 (43.4) |
| 2017 | \$8.91 | 110.6 (52.8) |
| 2018 | \$12.04 | 139.3 (58.5) |
| 2019 | \$14.61 | 167.1 (61.0) |

Σχήμα 1.1: Οι δαπάνες σε δισεκατομμύρια δολάρια για περιεχόμενο και οι συνδρομητές επί πληρωμή του Netflix σε εκατομμύρια (Joo, 2020).

Σύμφωνα με σχετική μελέτη, στη σύγχρονη εποχή η βιομηχανία της ψυχαγωγίας έχει κατακλυστεί από την παραγωγή ταινιών, η οποία καταλαμβάνει και το μεγαλύτερο μέρος της πρώτης. Για την ακρίβεια, ετησίως διατίθενται σε παγκόσμιο επίπεδο ένα σύνολο της τάξης των 4500 ταινιών. Το σύνολο αυτό ισοδυναμεί με 9000 ώρες περιεχόμενου (Huang & Wang, 2012).

Όπως μπορεί να γίνει αντιληπτό, πρόκειται για έναν πολύ μεγάλο αριθμό ταινιών, οι οποίες πρέπει να ταξινομηθούν σε ένα ή και περισσότερα είδη, ώστε να προσεγγίσουν το επιθυμητό κοινό. Μια τέτοια ταξινόμηση τείνει να πραγματοποιείται με τη συμμετοχή του σχετικού ανθρώπινου δυναμικού, με έναν πιο παραδοσιακό και καθόλου δυναμικό και αυτοματοποιημένο τρόπο (Huang & Wang, 2012). Αυτό έχει ως αποτέλεσμα να κοστίζει τόσο οικονομικά όσο και χρονικά και λαμβάνοντας υπόψη τον μεγάλο αριθμό αποτελεί έναν πολύ μεγάλο φόρτο εργασίας.

Συνεπώς, η αυτόματη αναγνώριση του είδους μιας ταινίας κερδίζει έδαφος. Η αυτοματοποίηση αυτής της διαδικασίας, όπως και πολλών άλλων, μπορεί να πραγματοποιηθεί με τη χρήση μεθόδων και τεχνικών που σχετίζονται με τον τομέα της Τεχνητής Νοημοσύνης (TN), σε συνδυασμό με τη χρήση διαφόρων στοιχείων που αφορούν την ταινία. Τέτοια στοιχεία μπορεί να αποτελούν οι αφίσες, τα clips ταινιών και κυρίως τα trailers τους.

Εκτός τούτου, η αυτόματη αναγνώριση μπορεί να ωφελήσει και τις συνδρομητικές υπηρεσίες streaming, όπως το προαναφερόμενο Netflix, οι οποίες χρησιμοποιούν συστήματα συστάσεων. Μιας και τα συστήματα αυτά δεν αποτελούν το επίκεντρο μελέτης της παρούσας εργασίας, αρκεί να ειπωθεί πως προτείνουν νέα στοιχεία (στην προκειμένη περίπτωση ταινίες) βάσει των προηγούμενων προτιμήσεων και αναζητήσεων του χρήστη (στην προκειμένη περίπτωση βάσει των ταινιών που παρακολουθούν).

Έτσι, η ορθή αναγνώριση και ταξινόμηση του είδους μιας ταινίας μπορεί ακόμα να επιφέρει ορθότερες και περισσότερο εξατομικευμένες προτάσεις στους χρήστες των πλατφορμών και των συνδρομητικών υπηρεσιών που διαθέτουν συστήματα συστάσεων, κάνοντας την υπόθεση πως σε έναν χρήστη αρέσουν οι ταινίες που ανήκουν στο ίδιο ή σε παρόμοιο είδος.

1.1 Αντικείμενο Εργασίας

Με δεδομένη τη σημασία των ταινιών στην καθημερινότητα και το μέγεθος της βιομηχανίας του κινηματογράφου, η όσο το δυνατόν πιο αντιπροσωπευτική ταξινόμηση των ταινιών στο αντίστοιχο είδος μπορεί να προσφέρει μεγαλύτερη ικανοποίηση στους θεατές, ενώ μπορεί να αποτελέσει τη βάση για ανάπτυξη θεωριών, γύρω από την επιτυχία και τα κέρδη μιας νέας ταινίας.

Κατά συνέπεια, στην παρούσα εργασία αναπτύσσεται και αξιολογείται μία μέθοδος αυτόματης αναγνώρισης του είδους μιας ταινίας, με βάση τα οπτικά χαρακτηριστικά του trailer που της αντιστοιχεί και, για την ακρίβεια, με βάση την ανίχνευση αντικειμένων. Πιο λεπτομερώς, ένα

προωθητικό προϊόν μιας ταινίας, το οποίο είναι και το πιο διαδεδομένο είναι το trailer. Σε ένα trailer προβάλλεται πλήθος σκηνών και διαλόγων της ταινίας, με σκοπό την προσέλκυση των θεατών. Κάθε σκηνή σε αυτό περιλαμβάνει μια ποικιλία αντικειμένων που σχετίζεται με την πλοκή και το επίκεντρο της ταινίας, υποδηλώνοντας το τί αναμένεται να παρακολουθήσει ο θεατής.

Προκύπτει, έτσι, εύλογα το ερώτημα του ποια και πόσα αντικείμενα που απεικονίζονται στις σκηνές ενός trailer μπορούν να αποτελέσουν χαρακτηριστικά αντικείμενα για την αναγνώριση του αντίστοιχου είδους. Για την επίλυση αυτού του προβλήματος, αρχικά απαιτείται η συλλογή ενός κατάλληλου συνόλου δεδομένων και συγκεκριμένα trailers, τα οποία μπορούν να χρησιμοποιηθούν, ώστε να αναγνωρίζεται αυτόματα το είδος μίας ταινίας. Έτσι, συλλέγονται συνολικά 209 trailer από ταινίες μέσω του YouTube και γίνεται λήψη των στοιχείων τους από την πλατφόρμα IMDb.

Για να γίνει αυτό, απαιτείται μία απαραίτητη προεπεξεργασία, όπου γίνεται ανάλυση κάθε trailer από frame σε frame, χρησιμοποιώντας τεχνικές αναγνώρισης αντικειμένων (object detection) για την καταγραφή του πλήθους εμφάνισης κάθε αντικειμένου σε αυτό. Στα πλαίσια της παρούσας εργασίας, επιλέγεται η τεχνική αναγνώρισης αντικειμένων You Only Look Once (YOLO), καθώς παρέχει άμεση απόκριση, που είναι κατάλληλη για εφαρμογή σε βίντεο.

Χρησιμοποιώντας τέτοιου είδους τεχνικές αναγνώρισης αντικειμένων, προκύπτει ένα συνολικό πλήθος εμφανίσεων κάθε αντικειμένου, η αναγνώριση του οποίου υποστηρίζεται από την εκάστοτε τεχνική, το οποίο συνιστά ένα νέο σύνολο δεδομένων. Με τη σειρά του, αυτό τροφοδοτεί την εκπαίδευση ευφυών αλγορίθμων, με στόχο την ταξινόμηση των ειδών μίας ταινίας. Ως πρόβλημα ταξινόμησης, το συγκεκριμένο είναι αρκετά σύνθετο, αφού μία ταινία μπορεί να ανήκει σε πολλά είδη, οπότε και συνιστά ένα multi-label πρόβλημα ταξινόμησης. Το πρόβλημα αυτό επιλύεται κάνοντας μετατροπή αυτού σε μία σειρά από binary προβλήματα, όπως είναι σύνηθες για την κατηγορία αυτή των προβλημάτων.

Για τη σύνθεση αυτού του συστήματος επεξεργασίας βίντεο με τελικό στόχο την αναγνώριση του είδους μίας ταινίας, γίνεται εφαρμογή μίας πληθώρας αλγορίθμων με τους οποίους πραγματοποιείται εκτενής πειραματισμός για την κατάλληλη παραμετροποίησή τους. Ιδιαίτερα σε σχέση με τους αλγορίθμους ταξινόμησης, γίνεται πειραματισμός με μεγάλη ποικιλία τεχνικών, οι οποίες συγκρίνονται και με απλούς ευρετικούς κανόνες με στόχο την αξιολόγηση της αποτελεσματικότητάς τους.

Κατά την αξιολόγηση της μεθοδολογίας, διερευνώνται και αξιολογούνται διάφορες μετρικές, τόσο για την αξιολόγηση της απόδοσης στο σύνθετο πρόβλημα multi-label ταξινόμησης, όσο και στα επιμέρους binary προβλήματα. Ακόμη, εξάγονται και αναλύονται τα κυριότερα αντικείμενα που προκύπτουν βάσει των αλγορίθμων, τα οποία επιβεβαιώνονται εν μέρει και από τη βιβλιογραφία.

1.2 Δομή Εργασίας

Στο δεύτερο κεφάλαιο αυτής της εργασίας γίνεται μία λεπτομερής ανασκόπηση των ειδών που χαρακτηρίζουν μία ταινία, των συσχετίσεων μεταξύ τους, αλλά και των διαφορετικών προσεγγίσεων που εντοπίζονται στη βιβλιογραφία για την αυτόματη αναγνώριση του είδους μιας ταινίας. Με τον τρόπο αυτόν, ο αναγνώστης είναι σε θέση να κατανοήσει πλήρως ποια χαρακτηριστικά διέπουν κάθε είδος και πώς έχει προσεγγιστεί στη βιβλιογραφία η αυτόματη αναγνώριση του είδους μιας ταινίας.

Στο τρίτο κεφάλαιο αναλύεται εκτενώς η μεθοδολογία που αναπτύσσεται στα πλαίσια της παρούσας εργασίας, συμπεριλαμβανομένης της διαδικασίας συλλογής του συνόλου δεδομένων των ταινιών και των αντίστοιχων trailers που αξιοποιούνται, της εξαγωγής και προεπεξεργασίας των απαραίτητων χαρακτηριστικών, της διαδικασίας ανίχνευσης αντικειμένων και, τέλος, των ευφύων αλγορίθμων ταξινόμησης με τους οποίους διεξάγεται πειραματισμός. Συμπερασματικά, το κεφάλαιο αναλύει με πλήρη λεπτομέρεια τις διαδικασίες που ακολουθήθηκαν, ώστε να προσφέρει πλήρη κατανόηση της μεθοδολογίας στον αναγνώστη.

Στο τέταρτο κεφάλαιο, παρουσιάζονται λεπτομέρειες σχετικά με τη διαδικασία πειραματισμού για την αξιολόγηση της μεθοδολογίας και των χρησιμοποιηθέντων αλγορίθμων, καθώς και τα αποτελέσματα που προέκυψαν από αυτή, χρησιμοποιώντας σχετικές μετρικές απόδοσης, ώστε να φανεί η καταλληλότητα των διαδικασιών και των τεχνικών που ακολουθήθηκαν. Ακόμη, γίνεται ανάλυση των κυριότερων αντικειμένων που ανιχνεύονται σε κάθε είδος ταινίας και μελετάται το κατά πόσο αυτά ανταποκρίνονται στην πραγματικότητα. Επιπλέον, γίνεται σύγκριση με τη βιβλιογραφία τόσο των ποσοτικών αποτελεσμάτων, όσο και των κυριότερων χαρακτηριστικών μάθησης, αλλά και γενικότερα του συνόλου της μεθοδολογίας.

Τέλος, στο πέμπτο κεφάλαιο γίνεται ένας απολογισμός όσων επιτεύχθηκαν στα πλαίσια της εργασίας εξάγοντας κάποια συμπεράσματα και εξετάζονται πιθανές μελλοντικές κατευθύνσεις για τη βελτίωση, τόσο της μεθοδολογίας, όσο και των αποτελεσμάτων, με στόχο την αυτόματη αναγνώριση και ταξινόμηση τους είδους μιας ταινίας.

2. Είδη Ταινιών

Μια κινηματογραφική παραγωγή μπορεί να κατηγοριοποιηθεί ως προς το κινηματογραφικό είδος, αλλά και ως προς το είδος της ταινίας. Ξεκινώντας από το πρώτο, μπορεί να πραγματοποιηθεί ο ακόλουθος διαχωρισμός. Υπάρχουν οι ταινίες αφήγησης, γνωστές και ως ταινίες μυθοπλασίας, οι οποίες επικεντρώνονται σε κάποια ιστορία που αποτελεί αποτέλεσμα μυθοπλασίας, αλλά ορισμένες φορές μπορεί να πηγάζει και από κάποιο πραγματικό γεγονός, όπως αυτό έχει αφηγηθεί, ενισχύοντας το στοιχείο του ρεαλισμού (Τάγκα, 2014).

Επιπλέον, ένα ακόμα κινηματογραφικό είδος είναι οι ταινίες ντοκιμαντέρ, στις οποίες επικρατεί το στοιχείο της αληθοφάνειας, ελλείπει οποιοδήποτε στοιχείου που αποτελεί δημιούργημα φαντασίας και μυθοπλασίας. Σε αντίθεση, λοιπόν, με τις ταινίες μυθοπλασίας, στα ντοκιμαντέρ δεν υπάρχουν διάλογοι μεταξύ ηθοποιών, μέσω των οποίων εξελίσσεται μια ιστορία, αλλά αποτυπώνεται η ισχύουσα πραγματικότητα, με στόχο να διατηρηθούν και να εμπλουτιστούν ιστορικά στοιχεία. Αξίζει να σημειωθεί πως το κόστος παραγωγής ενός ντοκιμαντέρ είθισται να είναι κατά πολύ μικρότερο από αυτό των ταινιών μυθοπλασίας (Τάγκα, 2014).

Εν συνεχεία, υπάρχουν τα animations, τα οποία αποτελούνται από σειρές εικόνων δύο ή τριών διαστάσεων που τοποθετούνται με τέτοιο τρόπο, ώστε να δίνουν την αίσθηση της κίνησης, φαινόμενο που αποτελεί μια οφθαλμαπάτη. Η ψευδαίσθηση αυτή οφείλεται στο γεγονός πως μια εικόνα εμφανίζεται για κλάσματα του δευτερολέπτου και ακολουθείται από την επόμενη. Αυτό το κινηματογραφικό είδος είναι ευρέως διαδεδομένο σε παιδικές ηλικίες, ενώ το κόστος εξαρτάται από το πόσο ανθρωπομορφικοί είναι οι χαρακτήρες (Τάγκα, 2014).

Όλες οι παραγωγές, ανεξαρτήτως κινηματογραφικού είδους και είδους ταινίας, προκειμένου να θεωρηθούν επιτυχημένες και να προσελκύσουν το κοινό, οφείλουν να προκαλούν ένα αίσθημα οικειότητας στον θεατή και συνάμα να χαρακτηρίζονται από μία πρωτοτυπία, ώστε να τον συνεπάρουν (Schatz, 2013). Επιπλέον, η παραγωγή νέων ταινιών επηρεάζεται από την αποτελεσματικότητα και την απήχηση προηγούμενων παραγωγών, αλλά στοχεύει και στην προβολή πιο καινοτόμων στοιχείων (Σκοπετέας, 2015).

Η κατηγοριοποίηση των κινηματογραφικών ταινιών σε είδη δεν στηρίζεται τόσο σε κάποιο θεωρητικό υπόβαθρο, αλλά ορίζεται βάσει των προσδοκιών και των απόψεων των θεατών σχετικά με τη ροή και την κατάληξη μίας ταινίας. Ακόμη, σημαντικό ρόλο στο χαρακτηρισμό του είδους μιας ταινίας παίζει η παραγωγή, η προώθηση και η διανομή των ταινιών (Neale, 1990).

Αξίζει να αποσαφηνιστεί, βέβαια, πως το είδος, βάσει της αρχαίας ελληνικής, αποτελεί μια ορολογία για οποιαδήποτε κατηγορία λογοτεχνίας ή άλλων μορφών τέχνης ή ψυχαγωγίας (όπως για παράδειγμα η μουσική) γραπτής ή προφορικής, ηχητικής ή οπτικής, βασισμένης σε κάποιο σύνολο στυλιστικών κριτηρίων.

Παράλληλα, όλα τα είδη ταινιών αποτυπώνουν τη κοινωνία και τα δρώμενα της αντίστοιχης εποχής και για τον λόγο αυτόν κάποια εξ αυτών γνωρίζουν μεγαλύτερη επιτυχία σε συγκεκριμένα κοινωνικά σύνολα και γεωγραφικές περιοχές (Σκοπετέας, 2015). Επί παραδείγματι, στην Αμερική γνωρίζουν μεγάλη άνθηση ταινίες με σούπερ ήρωες που σώζουν άτομα που το έχουν ανάγκη, μια κεντρική ιδέα όχι και τόσο συμβατή με την ελληνική κουλτούρα.

Σε γενικότερα πλαίσια, το είδος αποτελεί τον πυλώνα κατηγοριοποίησης μιας ταινίας και δηλώνει ομοιότητες στα αφηγηματικά στοιχεία, τα οποία χαρακτηρίζουν μια ταινία. Είθισται κάθε είδος να διέπεται από κάποιους κανόνες που ένας σκηνοθέτης οφείλει να ακολουθήσει. Οι κανόνες μπορεί να αφορούν στον φωτισμό, τα ηχητικά εφέ, την κίνηση της κάμερας, τα οποία με τη σειρά τους προκαλούν την αντίστοιχη ατμόσφαιρα και αντιδράσεις στους θεατές, μεταφέροντάς τους τις απαιτούμενες πληροφορίες και μηνύματα (Jain & Jadon, 2009).

Παρά το γεγονός πως είναι στην ευχέρεια του εκάστοτε σκηνοθέτη το πώς θα αξιοποιήσει τους προαναφερόμενους κανόνες, οι ταινίες του ίδιου είδους, όπως αναμένεται εξάλλου, έχουν πολλά κοινά στοιχεία (Αγιζον, 1991). Κατά συνέπεια, κάθε είδος αναγνωρίζεται από κάποια συγκεκριμένα μοτίβα και τεχνικές και κατ' επέκτασιν, μια ταινία μπορεί να ξεχωρίσει σε σύγκριση με άλλες, από την τοποθεσία, τη συναισθηματική επιβάρυνση που προκαλεί και από τον τρόπο που παρουσιάζεται με τη χρήση εξειδικευμένου εξοπλισμού (Jain & Jadon, 2009).

Παρά ταύτα, η συγκεντρωτική παρουσίαση των ειδών των εκάστοτε ταινιών δεν μπορεί να επιτευχθεί πλήρως, καθώς δεν έχουν αποσαφηνιστεί πλήρως η ποιότητα και η ποσότητα αυτών (McKee, 1999). Εν συνεχεία, ακολουθεί μια ανάλυση των δημοφιλέστερων ειδών ταινιών παγκοσμίως.

2.1 Περιγραφή Ειδών Ταινιών

Μεταξύ των πιο δημοφιλών ειδών ταινιών συγκαταλέγονται τα ακόλουθα:

- οι Γουέστερν ταινίες
- οι ταινίες Πολέμου
- οι ταινίες Τρόμου
- οι Βιογραφίες
- οι ταινίες Επιστημονικής Φαντασίας
- οι Περιπέτειες και οι ταινίες Δράσης
- οι Επικές ταινίες
- οι Κωμωδίες
- τα Δράματα
- οι ταινίες Thriller κ.ά.

Αξίζει να αναφερθεί πως πολλά είδη συγχέονται μεταξύ τους και συνεργατικά χαρακτηρίζουν μια ταινία. Για παράδειγμα, μια ταινία μπορεί να χαρακτηριστεί από κοινού ως ταινία τρόμου και thriller ταυτόχρονα. Ιδιαίτερα στη σύγχρονη βιομηχανία κινηματογραφικών παραγωγών συναντάται η παραγωγή ταινιών που συνδυάζουν στοιχεία από πολλά είδη, με σκοπό την

μεγαλύτερη δυνατή απήχηση. Τρανό παράδειγμα αποτελεί η ταινία “Star Wars”, η οποία περιλαμβάνει στοιχεία από πολυάριθμα είδη, μεταξύ των οποίων επιστημονική φαντασία, ταινίες πολέμου και Γουέστερν¹.

Ξεκινώντας από τις Γουέστερν ταινίες, πρόκειται για ταινίες που διαδραματίζονται στην Άγρια Δύση, στις οποίες ο πρωταγωνιστικός ρόλος αφορά ένα άτομο με μοναχική συμπεριφορά και ήθος, το οποίο καλείται να αντιμετωπίσει περισσότερο φιλοχρήματα άτομα που χαρακτηρίζονται από αηθικισμό, μέσω μαχών και εμπλοκών. Ομοίως στις ταινίες πολέμου, πρωταγωνιστούν οι βίαιες μάχες, με τον βασικό χαρακτήρα να προσπαθεί να επιβιώσει και να προάγει τις αξίες και τα ιδανικά που βρίσκονται σε κίνδυνο. Πολλές φορές αποτυπώνονται και πολιτικές στάσεις και η ζωή του άμαχου πληθυσμού (Σκοπετέας, 2015).

Από την άλλη, στις ταινίες τρόμου είθισται ο πρωταγωνιστής να είναι θύμα, να οδηγείται από τους φόβους του και να απειλείται από κάποια υπερφυσική δύναμη που αφορά κάποιο κοινωνικό, υπερφυσικό ή τεχνολογικό δημιούργημα. Συχνά, τα πρόσωπα παρουσιάζουν κάποια ιδιαίτερα χαρίσματα, ενώ η θρησκεία συντελεί ως μέσο για την αντιμετώπιση του κακού. Οι βιογραφίες, πάλι, μεταφέρουν πραγματικά στοιχεία από την καθημερινότητα και τα επιτεύγματα γνωστών ατόμων, που έχουν αφήσει το στίγμα τους στην ανθρωπότητα (Σκοπετέας, 2015).

Σε αντίθεση με το ρεαλισμό των βιογραφιών, οι ταινίες επιστημονικής φαντασίας παρουσιάζουν την ανθρωπότητα που κινδυνεύει από κάποιο ακραίο τεχνολογικό κατασκεύασμα ή από μια εξωγήινη/εξωπραγματική απειλή. Οι άνθρωποι συνεργάζονται και συμμαχούν ώστε να υπερνικήσουν την απειλή, προάγοντας την ελπίδα και την αλληλεγγύη (Σκοπετέας, 2015). Αναφορικά με τις ταινίες περιπέτειας και δράσης, δίνεται ιδιαίτερη προσοχή στην πλοκή που δεν μένει καθόλου στάσιμη, με τον πρωταγωνιστή να έχει συνήθως το ρόλο του σωτήρα έναντι ενός επικίνδυνου και παντοδύναμου προσώπου (Σκοπετέας, 2015).

Οι επικές ταινίες μοιάζουν με αυτές της περιπέτειας και της δράσης, καθώς ο πρωταγωνιστής είναι χαρισματικός και δεν διστάζει να θυσιάσει, προκειμένου να υπερνικήσει έναν ισχυρό αντίπαλο, ο οποίος συνήθως σχετίζεται με κάποιο ιστορικό γεγονός. Η προσωπική προσπάθεια του πρωταγωνιστή συγχέεται με αυτή του κοινωνικού συνόλου. Επιπλέον, οι κωμωδίες παρουσιάζουν καταστάσεις με κωμικά στοιχεία, ώστε να προκαλέσουν γέλιο και ευχαρίστηση στο κοινό. Ωστόσο, οι καταστάσεις αυτές ελλείπει των κωμικών στοιχείων, θα μπορούσαν να προκαλέσουν μελαγχολία και στεναχώρια (Σκοπετέας, 2015).

Ως προς τα δράματα, η πλοκή επικεντρώνεται στην προσωπική ζωή του πρωταγωνιστή, τη ψυχική του οδύνη και τις συναισθηματικές του συγκρούσεις (Banham & Brandon, 1995). Τέλος, τα thriller τείνουν να προκαλούν έντονα συναισθήματα, και κυρίως αυτά της φρίκης και της έξαψης, αποσκοπώντας στο να “κόψουν την ανάσα” του θεατή, εντείνοντας την αγωνία του σε μια ζοφερή ατμόσφαιρα. Εξαιτίας των συναισθημάτων που δημιουργεί, πολλές φορές

¹ https://en.wikipedia.org/wiki/Star_Wars

δύσκολα διακρίνεται η διαφορά των thriller με άλλα είδη ταινιών, όπως ταινίες τρόμου, περιπέτειας, κ.ο.κ.

Όπως είναι αναμενόμενο, κάποια είδη είναι προτιμότερα από κάποια άλλα στο ευρύτερο κοινό. Κατά την περίοδο Νοεμβρίου-Δεκεμβρίου του 2018 έλαβε χώρα μια online ψηφοφορία που ήταν ανοιχτή για τέσσερις ημέρες, στις Ηνωμένες Πολιτείες Αμερικής (ΗΠΑ), στην οποία συμμετείχαν 2200 άτομα, άνδρες και γυναίκες, ως προς τα είδη ταινιών που προτιμούν περισσότερο².

Σύμφωνα με τα αποτελέσματα, τα υψηλότερα ποσοστά συγκέντρωσαν οι κωμωδίες, καθώς τις προτιμά το 91% των γυναικών και το 90% των ανδρών αντίστοιχα. Ακολουθούν οι ταινίες που χαρακτηρίζονται ως δράματα, με 89% στις γυναίκες και 80% στους άνδρες, κι έπειτα οι ταινίες περιπέτειας με 89% στις γυναίκες και 90% στους άνδρες.

Αναφορικά με τις ταινίες επιστημονικής φαντασίας, παρατηρήθηκε και μια μεγάλη διαφορά μεταξύ των δύο φύλων, καθώς τις προτιμά το 76% των ανδρών έναντι του 62% των γυναικών. Τη μικρότερη προτίμηση κατείχαν οι ταινίες τρόμου, οι οποίες συγκέντρωσαν το μικρότερο ποσοστό στις προτιμήσεις των γυναικών (47%) και από τα μικρότερα ποσοστά στους άνδρες (57%), με το μικρότερο ποσοστό στις προτιμήσεις αυτών να εντοπίζεται στις ρομαντικές ταινίες.

Παράλληλα, οι προτιμήσεις του κοινού, όπως έχει ήδη αναφερθεί, διαφέρει από τόπο σε τόπο. Το γεγονός αυτό επιβεβαιώνεται και σε σχετική έρευνα από το “New York Film Academy”, τα ευρήματα της οποίας παρουσιάζονται στον Πίνακα 2.1. (New York Film Academy, 2016). Βάσει αυτών, φαίνεται πως το κοινό στις περισσότερες χώρες και σε διαφορετικές ηπείρους προτιμά τις κωμωδίες, ενώ ακολουθούν οι ταινίες δράσης. Αξίζει, επίσης, να σημειωθεί πως και στην έρευνα αυτή φαίνεται πως οι ΗΠΑ προτιμούν τις κωμωδίες, αποτέλεσμα που έρχεται σε συμφωνία με την προαναφερόμενη ψηφοφορία.

Πίνακας 2.1: Το προτιμότερο είδος ταινίας ανά χώρα (New York Film Academy, 2016).

| Χώρα | Είδος Ταινίας που προτιμάται |
|--------------------|------------------------------|
| Βραζιλία | Ταινίες Δράσης |
| Αργεντινή | Κωμωδίες |
| ΗΠΑ/Μεξικό/Καναδάς | Κωμωδίες |
| Ιαπωνία | Ταινίες Φαντασίας |
| Γερμανία | Ταινίες Δράσης/Περιπέτειες |
| Ινδία | Περιπέτειες |

² <https://www.statista.com/statistics/254115/favorite-movie-genres-in-the-us/>

| | |
|--------------|---------------------------------|
| Ιταλία | Κωμωδίες |
| Γαλλία | Κωμωδίες |
| Σουηδία | Δράματα/Κωμωδίες/Ταινίες Δράσης |
| Αυστραλία | Ταινίες Δράσης |
| Νέα Ζηλανδία | Ταινίες Δράσης/Κωμωδίες |
| Αγγλία | Κωμωδίες |

2.2 Αυτόματη Αναγνώριση Ειδών Ταινιών

Η αυτόματη ανάλυση trailers, τα οποία ουσιαστικά αποτελούν βίντεο, είναι μια διαδικασία που κερδίζει ολοένα και περισσότερο έδαφος και προσφέρει πολλές ερευνητικές δυνατότητες, όπως αναγνώριση ταινιών, κατηγοριοποίηση ταινιών, αναγνώριση στοιχείων, ανάλυση περιβάλλοντος και πολλές άλλες εργασίες.

Μια εύλογη εφαρμογή των παραπάνω με μεγάλη χρησιμότητα, ιδιαίτερα σε γονείς, είναι η αναγνώριση της καταλληλότητας του περιεχομένου μιας ταινίας για ανήλικα άτομα. Πολλές φορές η αναγνώριση της καταλληλότητας του περιεχομένου είναι δύσκολη και για τους ανθρώπους που κατηγοριοποιούν την κάθε σκηνή χειρωνακτικά. Σχετική έρευνα επικεντρώνεται στο συγκεκριμένο πρόβλημα και με στόχο τον έλεγχο καταλληλότητας, χρησιμοποιούνται αλγόριθμοι βαθιάς μάθησης (Wehrmann et al., 2018).

2.2.1 Βιβλιογραφική Ανασκόπηση

Πλήθος ερευνητών ανά τον κόσμο έχει ασχοληθεί με την αυτόματη αναγνώριση τους είδους ή των ειδών μιας ταινίας. Από το 2009 ακόμη, οι Jain & Jadon σε σχετική τους δημοσίευση βασίζονται σε clips ταινιών, προκειμένου να τις κατατάξουν σε είδη (Jain & Jadon, 2009). Από τα clips διαχωρίζουν τα ηχητικά από τα οπτικά χαρακτηριστικά .

Για τα οπτικά χαρακτηριστικά ακολουθείται η διαδικασία της τμηματοποίησης (segmentation), από την οποία εξάγονται frames, ενώ για τα ηχητικά η διαδικασία δειγματοληψίας. Ακόμα, η ταξινόμηση πραγματοποιείται με τη χρήση ενός Feed Forward Multilayer Perceptron νευρωνικού δικτύου, με επιβλεπόμενο αλγόριθμο μάθησης τον Backpropagation για την εκπαίδευση αυτού.

Παρ' όλ' αυτά, ο διαχωρισμός σε οπτικά και ηχητικά χαρακτηριστικά για την εκπαίδευση του νευρωνικού δεν κρίθηκε αποδοτικός, γι' αυτό και οι Jain & Jadon πραγματοποιούν μια μίξη των χαρακτηριστικών, με στόχο την επίτευξη πιο ικανοποιητικών αποτελεσμάτων. Εν γένει, καταφέρνουν να πετύχουν την εν μέρει αποτελεσματική ταξινόμηση κάποιων clips σε συγκεκριμένα είδη: δράσης, κωμωδία, μιούζικαλ, δράμα και τρόμου. Παραδόξως, δεν συμπεριλαμβάνουν στην δημοσίευση αυτή καθόλου μετρικές αξιολόγησης, όπως precision,

accuracy, recall κ.ά., γεγονός που δεν εξυπηρετεί τους αναγνώστες και την πλήρη αξιολόγηση της έρευνάς τους.

Ακολουθούν οι Huang & Wang που υλοποιούν ένα σύστημα ταξινόμησης είδους ταινίας, χρησιμοποιώντας και αυτοί οπτικά και ηχητικά χαρακτηριστικά (Huang & Wang, 2012). Ξεκινώντας από τα πρώτα αναφέρουν πως αυτά μπορούν να μετρηθούν βάσει χρόνου, κάτι που υποδηλώνει το ρυθμό μιας ταινίας και βάσει χώρου, κάτι που σχετίζεται με το περιεχόμενο πολυμέσων (π.χ. εικόνα, γραφικά, ομιλία, κ.ά.).

Πιο αναλυτικά, και αναφορικά με τα οπτικά χαρακτηριστικά, ένα βίντεο κατόπιν σχετικής προεπεξεργασίας μπορεί να χωριστεί σε βίντεο shots, μέσω της τεχνικής “shot boundary detection”, ώστε να προκύψουν τα χρονικά χαρακτηριστικά. Εκ παραλλήλου, από τα βίντεο shots επιλέγονται κάποια key-frames από τα οποία εξάγονται τα χωρικά χαρακτηριστικά.

Από την άλλη, τα ηχητικά σήματα, όπως τονίζουν οι Huang & Wang, μπορούν να μετρηθούν ως προς τον χρόνο και τη συχνότητα. Ως προς τον χρόνο, είναι δυνατό τα ηχητικά σήματα με τη πάροδο αυτού να παρουσιάζουν κάποια ιδιαίτερα χαρακτηριστικά για την ένταση και τον ρυθμό τους. Ως προς τη συχνότητα, μέσω του μετασχηματισμού Fourier, καθίσταται εφικτός ο μετασχηματισμός του χρονικού πεδίου στην αντίστοιχη ζώνη συχνοτήτων.

Έτσι, οι Huang & Wang επιλέγουν τα πιο σχετικά χαρακτηριστικά (feature selection), προκειμένου να διευκολυνθεί η διαδικασία της ταξινόμησης, χρησιμοποιώντας κατάλληλους αλγορίθμους. Για τις ανάγκες της έρευνάς τους, συλλέγουν από το “Apple Movie Trailers website” trailers από 223 ταινίες ποικίλων ειδών, όπως δράση, κωμωδία, δράμα, thriller. Τα είδη έχουν προκύψει από το IMDb.

Στο σύνολο οι Huang & Wang καταλήγουν με 277 χαρακτηριστικά (τόσο οπτικά όσο και ηχητικά), τα οποία εξάγονται από τα trailers. Επιπλέον, χρησιμοποιούν για ταξινόμηση τα Support Vector Machines (SVMs), με τη βοήθεια των οποίων πετυχαίνουν 91.9% accuracy. Τέλος, συμπεραίνουν πως τα ηχητικά χαρακτηριστικά διευκολύνουν σε μεγαλύτερο βαθμό τον διαχωρισμό των ειδών των ταινιών σε σύγκριση με τα οπτικά.

Περνώντας σε νεότερες χρονολογικά μελέτες, οι Shambharkar & Doja εστιάζουν στην ταξινόμηση των βίντεο από trailers ταινιών σε είδη, βάσει των ανθρώπινων πράξεων που προβάλλονται, ανεξαρτήτως από το φόντο. Για τον σκοπό αυτόν, χρησιμοποιούν δύο σύνολα δεδομένων, εκ των οποίων το ένα περιέχει 474 trailers από το YouTube, τα οποία αφορούν κάποιες Αμερικάνικες ταινίες που αποτελούν παραγωγές μεταξύ του 2010 και 2014 (Shambharkar & Doja, 2020).

Αν και το εν λόγω σύνολο δεδομένων περιείχε δέκα συνολικά κλάσεις βάσει των ειδών των ταινιών, οι Shambharkar & Doja επικεντρώνονται μόνο σε πέντε (κωμωδία, ταινία δράσης, ταινία τρόμου, έγκλημα και ταινία επιστημονικής φαντασίας), με δεδομένο ότι πολλές ταινίες ανήκαν σε περισσότερες από μία κλάσεις και με στόχο τη μείωση της υπολογιστικής πολυπλοκότητας.

Η μέθοδος που ακολουθούν αποτελείται από τέσσερα στάδια: προεπεξεργασία, τμηματοποίηση, εξαγωγή χαρακτηριστικών και ταξινόμηση. Στο πρώτο στάδιο και εφόσον το βίντεο χωριστεί σε frames, πραγματοποιείται μια αλλαγή του μεγέθους αυτών και έπειτα οι εικόνες μετατρέπονται στην κλίμακα του γκρι (grayscale images). Παράλληλα, για την αφαίρεση του θορύβου χρησιμοποιείται το adaptive median φίλτρο.

Στο στάδιο της τμηματοποίησης, ένα κατώφλι εφαρμόζεται για την αφαίρεση του φόντου και σύμφωνα με τη τιμή του κατωφλιού εξάγεται ένα τμήμα του προσκηνίου. Έπειτα, στο στάδιο της εξαγωγής χαρακτηριστικών παρατηρούνται οπτικά χαρακτηριστικά, όπως τα χρώματα και οι υφές και χαρακτηριστικά κίνησης, τα οποία χρησιμοποιούνται για την εκπαίδευση του ταξινομητή.

Τέλος, οι Shambharkar & Doja χρησιμοποιούν μια δική τους προσέγγιση που στηρίζεται σε βαθιά νευρωνικά δίκτυα για την ταξινόμηση και συγκεκριμένα ονομάζεται “DCNN-DHO” (Deep Convolutional Neural Network-Deer Hunting Optimization). Μεταξύ των μετρικών αξιολόγησης που χρησιμοποιούν περιλαμβάνονται τα ακόλουθα: accuracy, precision, sensitivity, specificity, F-measure, false alarm rate and false discovery rate. Αξίζει να αναφερθεί πως οι Shambharkar & Doja πετυχαίνουν 95.23% accuracy.

Συνεχίζοντας με τους Simões et al., αυτοί προκειμένου να ταξινομήσουν το trailer μιας ταινίας στο κατάλληλο είδος, δημιουργούν εξ αρχής ένα σύνολο δεδομένων (Simões et al., 2016). Το συγκεκριμένο σύνολο δεδομένων είναι διαθέσιμο δημόσια με την ονομασία “ Labeled Movie Trailer Data (LMTD)”. Αποτελείται από περισσότερα από 3500 trailers ταινιών, με διάρκεια μεταξύ 60 και 200 δευτερολέπτων και ημερομηνία κυκλοφορίας από το 1974. Ακόμη, να αναφερθεί ότι τα είδη των ταινιών που προβάλλονται στα trailers είναι γνωστά.

Ωστόσο, παρατηρούν από τη βιβλιογραφία πως ένα κλασικό πρόβλημα που προκύπτει είναι ότι πολλές ταινίες ανήκουν σε περισσότερα από ένα είδη και κάποιος ερευνητής κρίνουν ως ορθή την ταξινόμηση μιας ταινίας, ακόμα και όταν ο αλγόριθμος πετύχει έστω και ένα από τα είδη στα οποία μπορεί να ανήκει η ταινία. Έτσι, αν μια ταινία ανήκει σε τρία είδη, έχει πιθανότητα της τάξης του 75% να ταξινομηθεί σωστά από έναν αλγόριθμο από καθαρή τύχη.

Για τον παραπάνω λόγο, οι Simões et al., από το αρχικό σύνολο δεδομένων, χρησιμοποιούν ένα υποσύνολο με 1067 trailers ταινιών που αρχικά ανήκουν σε 22 είδη. Εκείνοι επιλέγουν κάθε ταινία να ανήκει σε ένα μόνο από τέσσερα διαφορετικά είδη, συμπεριλαμβανομένων των ταινιών δράσης, τρόμου, κωμωδιών και δραμάτων που έχουν διατηρηθεί από τα 22 αρχικά είδη.

Ακολουθεί η διαδικασία της προεπεξεργασίας του συνόλου δεδομένων, με τον καθαρισμό αυτού, καθώς πολλά βίντεο των trailers περιέχουν δεδομένα, όπως μαύρα περιγράμματα, τα οποία δεν είναι απαραίτητα για την αναγνώριση του είδους της ταινίας. Ύστερα, γίνεται η διαδικασία αύξησης των δεδομένων (data augmentation), η οποία αποσκοπεί στην αποφυγή του overfitting και περιλαμβάνει την αλλαγή των μεγεθών των αρχικών δεδομένων.

Για την ταξινόμηση οι Simões et al. χρησιμοποιούν ένα Convolutional Neural Network (CNN) και, μάλιστα, για να πετύχουν πιο ακριβή αποτελέσματα, κάνουν χρήση ενός αλγορίθμου μη επιβλεπόμενης μάθησης. Μέσω του αλγορίθμου αυτού γίνεται μια επισημείωση των σκηνών (labeled data) που περιέχει ένα trailer και μια αντιστοίχιση (mapping) μεταξύ αυτών και των ήδη επισημειωμένων ως προς το είδος ταινιών του συνόλου δεδομένων. Αναφορικά με τα αποτελέσματα, η συγκεκριμένη μελέτη πετυχαίνει accuracy έως και 73.5%, ξεπερνώντας την απόδοση παρόμοιων ερευνών.

Εκ παραλλήλου, οι Shambharkar et al. εστιάζουν στην ταξινόμηση ενός trailer ταινίας σε κάποιο είδος, προκειμένου να εξυπηρετήσουν τον υποψήφιο θεατή, βοηθώντας τον να καταλάβει τι αναμένεται να παρακολουθήσει και ποια η καταλληλότητα του περιεχομένου (Shambharkar et al., 2020). Για τον σκοπό αυτόν, δημιουργούν ένα σύνολο δεδομένων, το οποίο αποτελείται από trailers που διατίθενται σε διάφορα κανάλια του YouTube.

Συνολικά, το σύνολο δεδομένων αποτελείται από 390 trailers ταινιών, για τα οποία έχουν διασφαλίσει να ανήκουν μόνο σε ένα είδος. Έτσι, καταλήγουν τα trailers των ταινιών να ανήκουν σε μία εκ των τεσσάρων ακόλουθων ειδών: δράσης, κωμωδία, τρόμου και ρομαντική.

Έπειτα, από κάθε trailer ξεχωριστά εξάγουν ένα frame για κάθε δευτερόλεπτο, χωρίς να λαμβάνουν υπόψη το πρώτο και τα τελευταία πέντε δευτερόλεπτα, αφού πρώτα και στα πλαίσια της προεπεξεργασίας των δεδομένων, έχουν χρησιμοποιήσει το λογισμικό FFmpeg, ώστε να πραγματοποιήσουν υποδειγματοληψία (downsampling).

Για την ταξινόμηση χρησιμοποιούν ένα 3D CNN μοντέλο, το οποίο λαμβάνει υπόψη τόσο τα χαρακτηριστικά που σχετίζονται με τον χώρο, όσο και αυτά που σχετίζονται με τον χρόνο και η αρχιτεκτονική του περιλαμβάνει 6 επίπεδα. Το γεγονός ότι εξετάζουν και τα χρονικά χαρακτηριστικά, οδηγεί σε υψηλότερη απόδοση των 3D CNNs, εν συγκρίσει με τα κλασσικά 2D CNNs, γι' αυτό και προτιμώνται στην εν λόγω μελέτη.

Οι Shambharkar et al. χωρίζουν το σύνολο δεδομένων σε training και test set, με το πρώτο να περιέχει 37460 frames και το δεύτερο 12846. Προκειμένου να ολοκληρωθεί η διαδικασία της εκπαίδευσης (training) κάνουν χρήση του Google Colaboratory, με την καθεμία από τις 60 συνολικά εποχές του νευρωνικού δικτύου να διαρκεί πέντε λεπτά. Σχετικά με τα αποτελέσματα, το μοντέλο τους επιτυγχάνει 84.12% accuracy.

Παρά ταύτα, η ταξινόμηση μιας ταινίας στο αντίστοιχο είδος, μπορεί να πραγματοποιηθεί και από την αφίσα της ταινίας, εκτός του trailer της. Μια σχετική έρευνα παρουσιάζουν οι Chu & Guo, οι οποίοι, κατασκευάζουν ένα σύνολο δεδομένων, συλλέγοντας αφίσες ταινιών από το IMDb σε διαφορετικές αναλύσεις (Chu & Guo, 2017). Από κάθε ταινία έχουν συλλέξει μία αφίσα, ενώ αξίζει να αναφερθεί πως οι ταινίες που έχουν επιλέξει είναι παραγωγές του Hollywood που έχουν λάβει χώρα μεταξύ 1980 και 2015.

Εκτός από τις αφίσες, έχουν συγκεντρώσει κι άλλες πληροφορίες για τις ταινίες, όπως το είδος τους, καταλήγοντας σε 23 διαφορετικά είδη. Στην προκειμένη περίπτωση, οι Chu & Guo λαμβάνουν υπόψη πως μια ταινία μπορεί να ανήκει σε περισσότερα από ένα είδη, και ως διαδικασία προεπεξεργασίας των δεδομένων χρησιμοποιούν την αύξηση αυτών. Επιπλέον, από το IMDb έχουν στη διάθεσή τους κι άλλες πληροφορίες, ουσιαστικά μεταδεδομένα για τις ταινίες, όπως οι βαθμολογίες της, ονόματα ηθοποιών και σκηνοθέτη, κ.ά., τα οποία, όμως, δεν εκμεταλλεύονται στην έρευνά τους.

Με τη βοήθεια ενός CNN ταξινομούν τις αφίσες σε είδη, με το CNN να αποτελείται από 7 convolutional επίπεδα. Εκτός από την οπτική αναπαράσταση της αφίσας για την εκπαίδευση του νευρωνικού δικτύου κάνουν χρήση και της ανίχνευσης αντικειμένων με τη μέθοδο YOLO, η οποία αναλύεται εκτενώς σε επόμενο Κεφάλαιο της παρούσας εργασίας.

Το γεγονός πως μια ταινία μπορεί να ανήκει σε παραπάνω από ένα είδος, επηρεάζει και τη διαδικασία της ταξινόμησής της. Κατά συνέπεια, οι Chu & Guo χρησιμοποιούν κάποια κατώφλια για τις πιθανότητες που έχουν καθοριστεί από το CNN, τα οποία με τη σειρά τους έχουν οριστεί από έναν αλγόριθμο grid αναζήτησης.

3. Μεθοδολογία Αναγνώρισης Είδους Ταινίας από Trailer

Σε αυτό το Κεφάλαιο, αναλύονται τα βήματα της μεθοδολογίας που αναπτύσσεται για την αναγνώριση του είδους μίας ταινίας με βάση το trailer της. Η Πτυχιακή εργασία που έδωσε έμπνευση για την ανάπτυξη της μεθοδολογίας είναι η <<Movie genre recognition>> (Alrowaili R. 2019). Για την ακρίβεια, περιγράφεται το αρχικό σύνολο δεδομένων με ταινίες, το οποίο αξιοποιείται για τη λήψη των πληροφοριών και των trailer ενός πλήθους ταινιών. Έπειτα, αναλύεται η διαδικασία προεπεξεργασίας της πληροφορίας αυτής, ώστε τελικά να προκύψουν τα χαρακτηριστικά και, κατ' επέκταση, το σύνολο δεδομένων που χρησιμοποιείται για την αναγνώριση του είδους των ταινιών.

Η κεντρική ιδέα πίσω από την προσέγγιση που ακολουθείται στα πλαίσια της παρούσας εργασίας είναι η εξής: Μία ταινία που ανήκει σε μία συγκεκριμένη κατηγορία είναι μάλλον σίγουρο ότι θα περιλαμβάνει συγκεκριμένα οπτικά στοιχεία, τα οποία ενδεχομένως να είναι μοναδικά για αυτό το είδος ταινιών. Ένα βασικό ερώτημα είναι: *Ακριβώς ποια είδη οπτικών στοιχείων παρέχουν, όντως, χρήσιμη πληροφορία που να συμβάλλει στην αυτόματη αναγνώριση του είδους μίας ταινίας;*

Ένα χαρακτηριστικό παράδειγμα, που ίσως δώσει μία κατεύθυνση στην απάντηση αυτού του ερωτήματος, είναι οι ταινίες τρόμου. Είναι βέβαιο πως τέτοιου είδους ταινίες θα διαδραματίζονται σε πιο σκοτεινά σκηνικά. Είναι εξίσου πιθανό να υπάρχουν απότομες μεταβάσεις μεταξύ των σκηνών με τρομακτικά ηχητικά εφέ, αν και εδώ δίνεται περισσότερο έμφαση στα οπτικά χαρακτηριστικά. Ακόμη, είναι μάλλον πιθανό να περιέχονται κάποια ιδιαίτερα αντικείμενα σε συγκεκριμένες σκηνές, όπως θα ήταν ένα μαχαίρι ή ένα όπλο σε μία σκηνή φόνου, τα οποία δεν θα παρουσιάζονταν σε κάποια είδη ταινίας, όπως οι κωμωδίες. Από την άλλη, όμως, θα ήταν αρκετά πιθανή η εμφάνιση τέτοιων αντικειμένων σε ταινίες μυστηρίου ή δράσης.

Με αυτή τη σκέψη, προκύπτει ένα από τα βασικά ερωτήματα που μελετά η παρούσα εργασία: *Πόσο σημαντικό ρόλο παίζει η εμφάνιση συγκεκριμένων αντικειμένων στις σκηνές μίας ταινίας για την αναγνώριση του είδους της;*

Ο αμέσως επόμενος προβληματισμός που προκύπτει είναι το πόσες και ποιες σκηνές είναι αντιπροσωπευτικές για μία ταινία. Εξάλλου, οι σύγχρονες ταινίες υψηλής ποιότητας περιέχουν σημαντικό όγκο πληροφορίας, ο οποίος δεν είναι εύκολα διαχειρίσιμος σε μεγάλη κλίμακα, δεδομένου ότι η επεξεργασία των βίντεο δεν είναι μία υπολογιστικά εύκολη διαδικασία. Επομένως, το να υποστεί επεξεργασία ολόκληρη η ταινία δεν είναι μία πρακτική λύση.

Μία συνετή προσέγγιση είναι η χρήση των trailers μίας ταινίας. Άλλωστε, τα trailers σχεδιάζονται με τέτοιον τρόπο, ώστε να προσελκύουν το ενδιαφέρον του θεατή και να του δίνουν μία όσο το δυνατόν πιο αντιπροσωπευτική προεπισκόπηση του περιεχομένου μίας ταινίας. Επομένως, είναι αρκετά πιθανό, αν υπάρχουν σκηνές που να σχετίζονται άμεσα με το

περιεχόμενο της ταινίας, αυτές να βρίσκονται στο trailer αυτής. Κατ' επέκταση, τα αντικείμενα που εμφανίζονται σε τέτοιες σκηνές, θα είναι μάλλον χαρακτηριστικά για την πλοκή αυτής, άρα και για το είδος της.

Συνδυαστικά, οι παραπάνω συλλογισμοί διαμορφώνουν πλήρως το αντικείμενο της παρούσας εργασίας, με κύριο στόχο την εξακρίβωση του αν τα αντικείμενα που εμφανίζονται κατά τη διάρκεια μιας ταινίας είναι αντιπροσωπευτικά για το είδος της. Για να επιτευχθεί αυτός, αναπτύσσεται η μεθοδολογία που παρουσιάζεται συνοπτικά στο Σχήμα 3.1.

Αρχικά, απαραίτητο είναι να γίνει λήψη μίας συλλογής δεδομένων που να περιλαμβάνει τουλάχιστον τα βίντεο από τα trailers ταινιών, καθώς και μεταδεδομένα για αυτές, συμπεριλαμβανομένου του είδους της ταινίας. Έπειτα, κάθε βίντεο θα υποστεί κατάλληλη προεπεξεργασία, ώστε να μειωθεί η υπολογιστική ισχύς που απαιτείται στα επόμενα βήματα. Στη συνέχεια, εφαρμόζεται μέθοδος αναγνώρισης αντικειμένων για κάθε frame, τα αποτελέσματα της οποίας καταγράφονται και αποτελούν ένα νέο σύνολο δεδομένων. Με βάση το τελευταίο, γίνεται εκπαίδευση αλγορίθμων TN, ώστε να εξαχθεί ένα μοντέλο που να είναι σε θέση να αναγνωρίζει αυτόματα το είδος μίας ταινίας με ικανοποιητική ακρίβεια. Στη συνέχεια, αναλύεται κάθε επιμέρους βήμα της μεθοδολογίας.



Σχήμα 3.1: Η μεθοδολογία που έχει ακολουθηθεί για την αυτόματη αναγνώριση του είδους μιας ταινίας.

3.1 Σύνολο Δεδομένων

Το αρχικό πρόβλημα που χρήζει αντιμετώπισης στη μεθοδολογία είναι να βρεθεί ένα σύνολο δεδομένων που να παρέχει με δομημένο τρόπο όλες τις απαραίτητες πληροφορίες. Σε αυτές συμπεριλαμβάνονται τα βασικά στοιχεία κάθε ταινίας, όπως είναι ο τίτλος της, τα είδη στα οποία αυτή ανήκει, καθώς και το trailer αυτής. Έχοντας αυτές τις πληροφορίες, μπορούν να λάβουν χώρα τα υπόλοιπα βήματα της μεθοδολογίας, ώστε τελικά να γίνει η εκπαίδευση ενός αλγορίθμου TN για να εξαχθεί ένα μοντέλο, το οποίο είναι σε θέση να αναγνωρίσει αυτόματα το είδος μιας ταινίας με βάση τα χαρακτηριστικά του trailer αυτής.

Για να γίνει με έναν συστηματικό τρόπο η λήψη της απαραίτητης πληροφορίας, επιλέγεται ένα από τα υπάρχοντα σύνολα δεδομένων που έχουν προταθεί στη βιβλιογραφία. Μιας και υπάρχει ένα πλήθος από διαθέσιμα σύνολα δεδομένων που αποσκοπούν στην αναγνώριση του είδους μιας ταινίας μέσω του trailer αυτής, γίνεται επιλογή του MovieNet³ (Huang et al., n.d.), καθώς είναι μία από τις πιο πρόσφατες προτάσεις που παρουσιάζεται στη βιβλιογραφία.

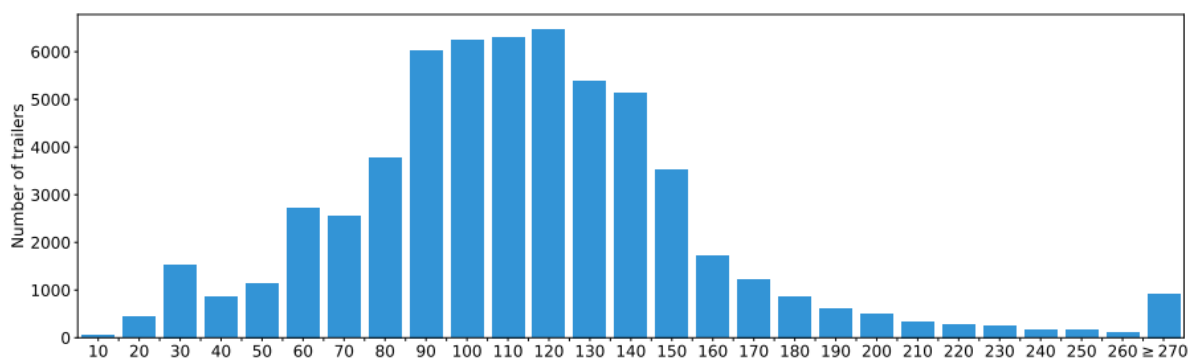
3.1.1 Περιγραφή Συνόλου Δεδομένων

Το MovieNet είναι ένα από τα πιο πρόσφατα και ολοκληρωμένα σύνολα δεδομένων που παρέχουν συγκεντρωτικά πληροφορίες που αφορούν σε ταινίες. Σε αυτό, περιέχονται διαφορετικά είδη πληροφορίας, όπως επισημειωμένα βίντεο και μεταδεδομένα, που εστιάζουν σε μία πληθώρα προβλημάτων, με επίκεντρο την υπολογιστική κατανόηση του περιεχομένου μιας ταινίας, συμπεριλαμβανομένων των στυλιστικών και σκηνοθετικών επιλογών, της πλοκής, μέχρι και της αναγνώρισης χαρακτήρων.

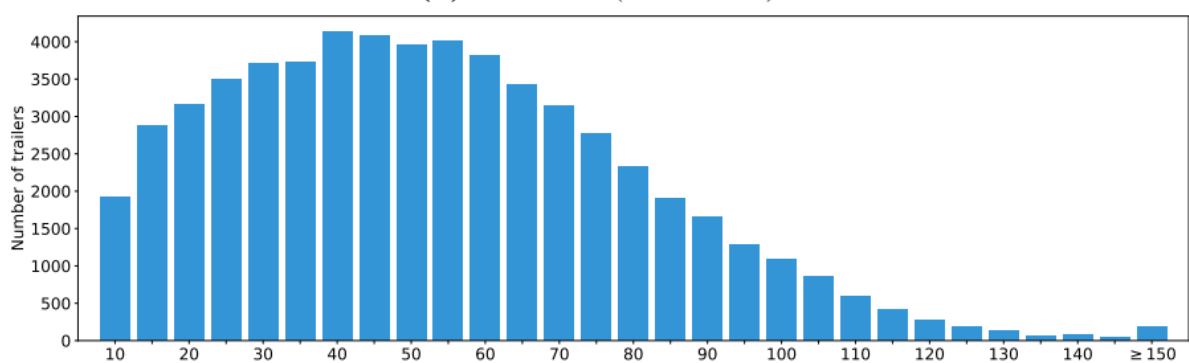
Μεταξύ των διαφόρων προβλημάτων, είναι και η αυτόματη αναγνώριση του είδους μιας ταινίας. Για τον σκοπό αυτόν, παρέχονται συνολικά πληροφορίες για περισσότερα από 60.000 trailers που αφορούν 33.000 ταινίες. Μερικά ενδεικτικά χαρακτηριστικά αυτών των trailers παρουσιάζονται στο Σχήμα 3.2. Τα στοιχεία που παρέχονται για αυτά τα trailers περιλαμβάνουν το αναγνωριστικό ID της ταινίας στο IMDb⁴, μία από τις δημοφιλέστερες ιστοσελίδες που αποσκοπούν στην πληροφόρηση σχετικά με ταινίες, η οποία διαθέτει μία από τις εκτενέστερες βάσεις δεδομένων. Επίσης, παρέχεται το αναγνωριστικό ID του trailer της ταινίας στο YouTube. Οι πληροφορίες αυτές είναι διαθέσιμες μέσω ενός αρχείου JSON.

³ <http://movienet.site/>

⁴ <https://www.imdb.com/>



(a) Duration (in seconds)



(b) Number of shots

Σχήμα 3.2: Η κατανομή της διάρκειας και των σκηνών στα trailers του MovieNet (Huang et al., n.d.).

Μέσω των δύο παρεχόμενων αναγνωριστικών, είναι δυνατή η ανάπτυξη εργαλείων για τη λήψη των απαραίτητων στοιχείων, ώστε να ακολουθήσει η αναγνώριση αντικειμένων στα trailers των ταινιών, με σκοπό την εξαγωγή των χαρακτηριστικών που θα συνθέσουν το σύνολο δεδομένων για την αυτόματη ταξινόμηση. Για την ακρίβεια, αναπτύσσεται πρόγραμμα σε Python 3.7, με στόχο τη λήψη των trailers και των απαραίτητων μεταδεδομένων (τίτλος και είδος ταινίας).

Σχετικά με τη λήψη των trailers, χρησιμοποιείται η βιβλιοθήκη pytube⁵, η οποία παρέχει μία προγραμματιστική διεπαφή για το YouTube, επιτρέποντας την αναζήτηση βίντεο μέσω URL, Επίσης, επιτρέπει τη λήψη βίντεο σε διαφορετικές ποιότητες και με διαφορετικές κωδικοποιήσεις. Σε αυτήν την περίπτωση, γίνεται λήψη όσων βίντεο είναι διαθέσιμα σε ανάλυση 1280x720p και μορφότυπο MP4. Τα βίντεο αυτά αποθηκεύονται σε συγκεκριμένο ευρετήριο, με βάση το όνομά τους.

Για τη διαχείριση των μεταδεδομένων, χρησιμοποιείται η βιβλιοθήκη IMDbPY⁶, η οποία παρέχει μία προγραμματιστική διεπαφή προς τη βάση δεδομένων του IMDb. Έτσι, διευκολύνει την προσπέλαση και λήψη των απαραίτητων μεταδεδομένων, τα οποία αποθηκεύονται σε αρχείο με μορφή Comma-Separated Values (CSV). Για την αντιστοίχιση μίας ταινίας σε αυτό

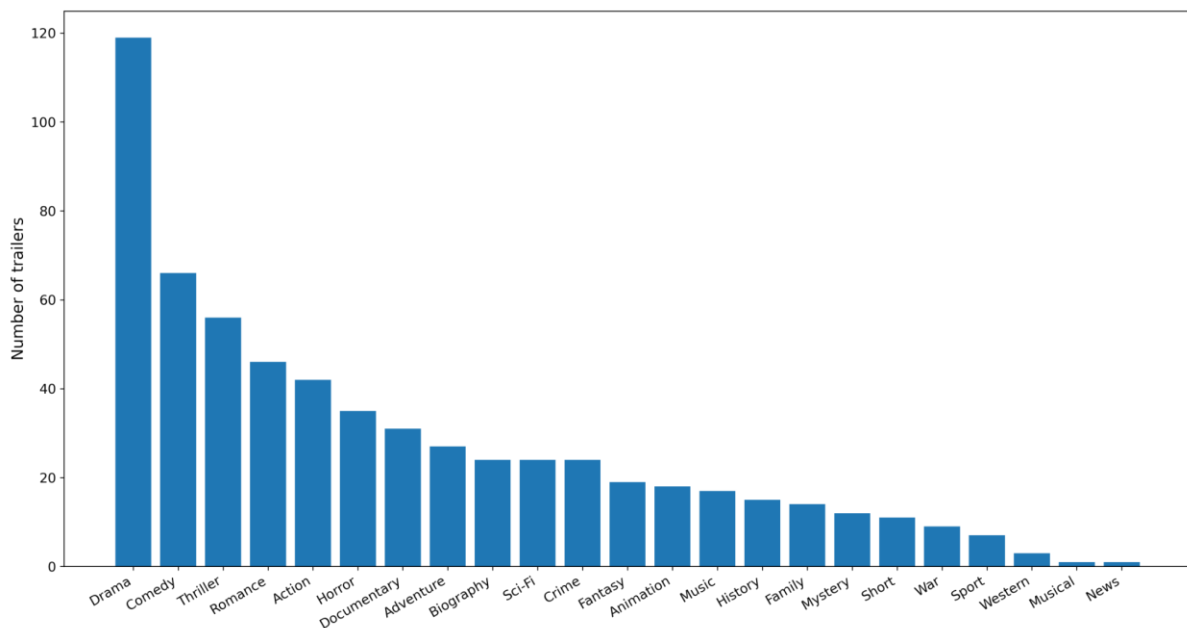
⁵ <https://pypi.org/project/pytube/>

⁶ <https://pypi.org/project/IMDbPY/>

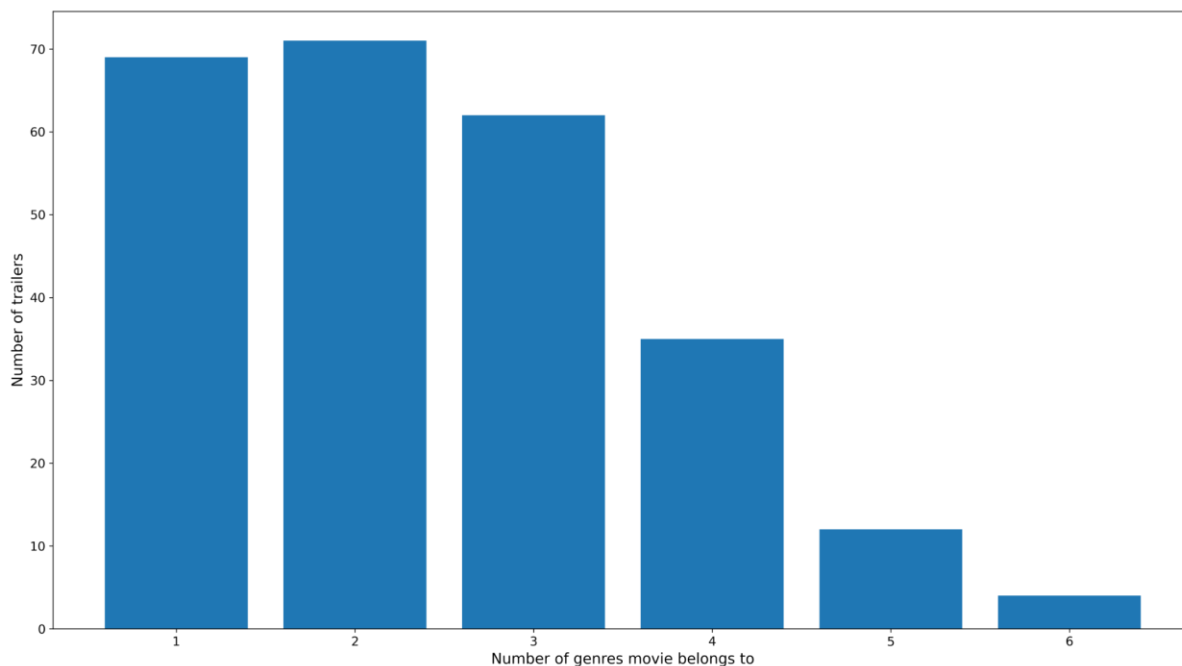
το αρχείο με το κατάλληλο trailer, αποθηκεύεται επίσης και η διαδρομή (path) προς το αποθηκευμένο βίντεο μαζί με τα υπόλοιπα μεταδεδομένα.

Σε αυτό το σημείο, αξίζει να σημειωθεί πως, σε περίπτωση που κάποια από τα παραπάνω χαρακτηριστικά δεν μπορούν να εντοπιστούν επιτυχώς, είτε διότι κάτι πιθανώς να έχει αλλάξει στα αναγνωριστικά και πλέον να μην αναγνωρίζονται από τις προηγούμενες ιστοσελίδες, είτε διότι δεν υπάρχει σε επαρκή ανάλυση το trailer, η ταινία παραλείπεται και τα στοιχεία της δεν αποθηκεύονται. Επίσης, το συγκεκριμένο πρόγραμμα έχει γραφεί με τρόπο τέτοιο ώστε, σε περίπτωση σφάλματος, να είναι σε θέση να συνεχίσει η λήψη των στοιχείων από το τελευταίο επιτυχημένο βίντεο. Έτσι, σε περίπτωση που συμβεί κάποιο λάθος, η διαδικασία δε χρειάζεται να ξεκινήσει από το μηδέν.

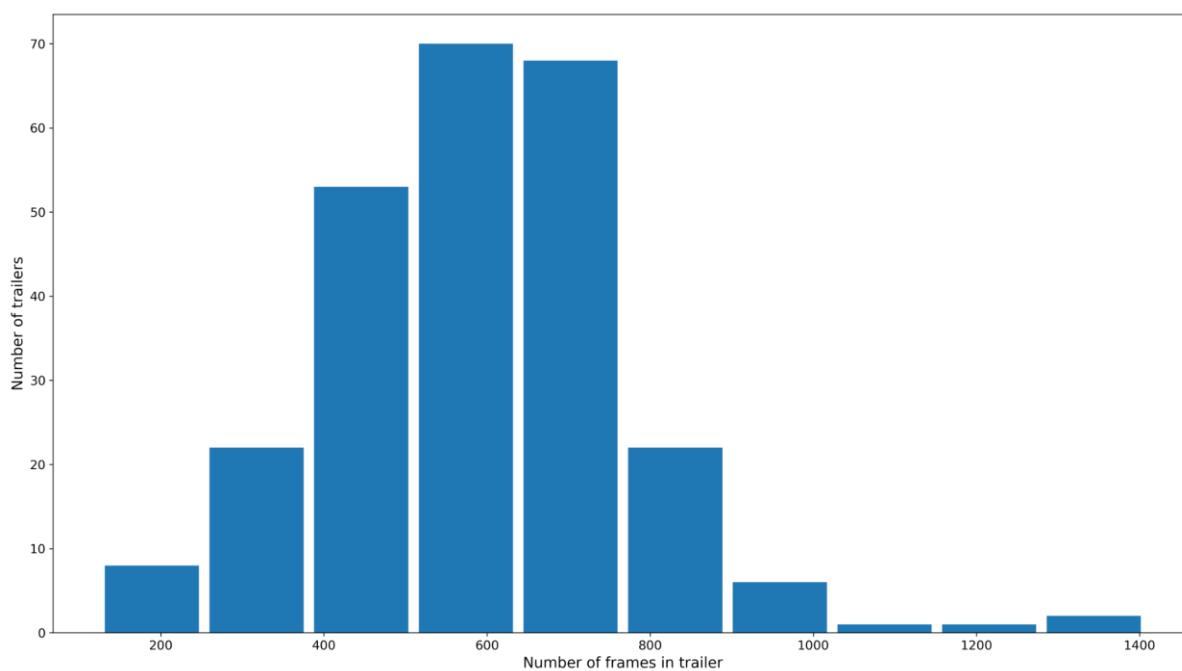
Χρησιμοποιώντας το πρόγραμμα αυτό, συνολικά έγινε λήψη 254 trailers ταινιών, οι οποίες συνολικά καταλαμβάνουν περισσότερο από 6.2 GB χώρο. Μερικά ενδιαφέροντα χαρακτηριστικά αναφορικά με τα είδη των ταινιών παρουσιάζονται στα Σχήματα 3.3 έως 3.5. Για την ακρίβεια, στο Σχήμα 3.3 παρουσιάζεται η συχνότητα εμφάνισης κάθε είδους ταινίας με τη μορφή ιστογράμματος. Στο Σχήμα 3.4 καταγράφεται η κατανομή του πλήθους των ειδών στα οποία κατατάσσονται οι ταινίες, ενώ στο Σχήμα 3.5 παρουσιάζεται η κατανομή του πλήθους των frames στα trailers με τη μορφή ιστογράμματος.



Σχήμα 3.3: Το πλήθος των ταινιών ανά κατηγορία στο αρχικό σύνολο δεδομένων που συλλέγεται.



Σχήμα 3.4: Το πλήθος των ταινιών συναρτήσει του πλήθους των κατηγοριών στις οποίες ανήκουν στο αρχικό σύνολο δεδομένων που συλλέγεται.



Σχήμα 3.5: Ιστόγραμμα του πλήθους των frames στις ταινίες που απαρτίζουν το αρχικό σύνολο δεδομένων.

3.1.2 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία των βίντεο είναι ένα αναπόσπαστο μέρος της διαδικασίας, αφού η επεξεργασία που θα ακολουθήσει θα εφαρμοστεί σε εκατοντάδες βίντεο, ώστε να πραγματοποιηθεί η αναγνώριση των αντικειμένων. Στόχος της προεπεξεργασίας είναι να μειωθεί τόσο η ανάλυση των βίντεο, όσο και το πλήθος των frames τους, επιτυγχάνοντας ταυτόχρονα ελάττωση της διαστατικότητας ως προς τον χώρο αλλά και τον χρόνο. Να

σημειωθεί πως το στάδιο της προεπεξεργασίας αποτελεί ακρογωνιαίο λίθο σε κάθε είδους πρόβλημα που σχετίζεται με την ανάλυση δεδομένων και περιλαμβάνει πλήθος τεχνικών (García et al., 2015).

Πιο συγκεκριμένα, για την πραγματοποίηση της προεπεξεργασίας, αναπτύσσεται πρόγραμμα σε Python 3.7 με τη βοήθεια της βιβλιοθήκης OpenCV⁷, το οποίο δέχεται ως είσοδο ένα βίντεο από το trailer μίας ταινίας και κρατάει κάθε frame αυτού. Έπειτα, επιλέγεται 1 frame ανά 5 του αρχικού βίντεο, το οποίο συρρικνώνεται σε μέγεθος, ώστε η ανάλυσή του να είναι 128x128. Έτσι, συνολικά επιτυγχάνεται 98.2% μείωση του αριθμού των pixels ανά frame, ενώ ταυτόχρονα μειώνεται το πλήθος των frames κατά 80%.

Συνδυαστικά, έπειτα από την προεπεξεργασία, μειώνεται δραστικά ο όγκος των βίντεο, επιτρέποντας την αποδοτικότερη αξιοποίησή τους στα επόμενα βήματα.

3.1.3 Ανίχνευση Αντικειμένων

Η αναγνώριση αντικειμένων (Liu et al., 2020) είναι μία σύνθετη διαδικασία, που έχει γνωρίσει μεγάλη άνθηση την τελευταία δεκαετία, εξαιτίας, εν μέρει, της ραγδαίας ανάπτυξης των αρχιτεκτονικών Βαθιών Νευρωνικών Δικτύων, ή Deep Neural Networks (DNNs). Συγκεκριμένα, οι τεχνικές αναγνώρισης αντικειμένων εστιάζουν στην ανίχνευση της εμφάνισης ενός αντικειμένου που ανήκει σε ένα (συνήθως μεγάλο) πλήθος προδιαγεγραμμένων κλάσεων σε μία εικόνα και, σε περίπτωση που υπάρχουν, είναι σύνθησης να προσδιορίζεται ακριβώς η θέση τους (Zou, 2019, Zou et al., 2019).

Είναι φυσικό αυτό να είναι ένα σύνθετο πρόβλημα, αφού υπάρχουν τουλάχιστον χιλιάδες διαφορετικά αντικείμενα σε ένα στιγμιότυπο του περιγύρου μας, με αποτέλεσμα η ερευνητική κοινότητα να εστιάζει συνήθως σε ένα εύρος “καλά δομημένων” κατηγοριών, της τάξης των δεκάδων ή εκατοντάδων αντικειμένων. Φυσικά, συνήθως προτιμώνται κατηγορίες που να είναι χρήσιμες σε πρακτικές εφαρμογές, όπως τα αυτοκίνητα, οι άνθρωποι, κ.ά (Kastrinaki et al., 2003).

Περνώντας, έτσι, στις περιοχές όπου η ανίχνευση αντικειμένων βρίσκει εφαρμογές, οι οποίες γενικά ποικίλουν και είναι πολλά υποσχόμενες, θα γίνει περεταίρω ανάλυση. Αρχικά, ένας σημαντικός τομέας είναι η μηχανική και ρομποτική όραση, η οποία μπορεί να χρησιμοποιηθεί για τον έλεγχο και την αυτοματοποίηση διαδικασιών, ιδιαίτερα σε βιομηχανικά περιβάλλοντα (Ekvall et al., 2007).

Μία επίσης ιδιαίτερα ανερχόμενη περιοχή είναι τα αυτόνομα οχήματα, στα οποία η ακριβής ανίχνευση των αντικειμένων είναι καθοριστικός παράγοντας της επιτυχίας τους, ιδιαίτερα σε ό,τι αφορά θέματα ασφάλειας. Ακόμη, παίζει σημαντικό ρόλο στην ανάκτηση εικόνων, καθώς και σε διάφορες εμπορικές συσκευές, όπως η έξυπνη παρακολούθηση μέσω κάμερας και πολλές άλλες (Janai et al., 2020).

⁷ <https://opencv.org/>

Συνεπώς, είναι εμφανές πως οι τεχνικές ανίχνευσης αντικειμένων είναι κρίσιμες και είναι σχεδόν βέβαιο πως θα παίζουν σημαντικό ρόλο σε τεχνολογικά και, κατ' επέκταση, κοινωνικά ζητήματα στο μέλλον. Για τον λόγο αυτόν, είναι απαραίτητο να παρουσιάζουν μεγάλη ακρίβεια στα αποτελέσματά τους. Ένα εξίσου σημαντικό χαρακτηριστικό είναι η ταχύτητα της τεχνικής, δεδομένου ότι κατά κύριο λόγο έχει ενδιαφέρον η ανίχνευση αντικειμένων σε βίντεο και όχι σε μία μεμονωμένη εικόνα (Karasulu & Korukoglu, 2013). Αν αναλογιστεί κανείς πως στα περισσότερα βίντεο, ένα δευτερόλεπτο αντιστοιχεί σε τουλάχιστον 24 frames και ότι για την πλειοψηφία των εφαρμογών, είναι απαραίτητη η ανίχνευση σε πραγματικό χρόνο, η ταχύτητα είναι ίσως το πιο κρίσιμο χαρακτηριστικό.

3.1.3.1. YOLO

Σε ακριβώς αυτήν την ανάγκη για άμεση απόκριση στην αναγνώριση αντικειμένων εστιάζει η μέθοδος YOLO (Redmon & Farhadi, 2018). Η μέθοδος αυτή, όπως και οι περισσότερες πρόσφατες στη βιβλιογραφία, αξιοποιεί DNN, του οποίου η αρχιτεκτονική βασίζεται στην πρόβλεψη των πλαισίων που προσδιορίζουν τη θέση εμφάνισης ενός αντικειμένου και την κατηγορία αυτού, χρησιμοποιώντας ως είσοδο μία εικόνα. Έτσι, για ένα βίντεο, χρησιμοποιείται το κάθε frame αυτού ως είσοδος προς το νευρωνικό δίκτυο.

Η αρχιτεκτονική του YOLO παρουσιάζει ενδιαφέρον, διότι αξιοποιεί ένα μεγάλο πλήθος από convolutional layers, συνιστώντας έτσι ένα CNN, σε αντίθεση με άλλες πιο περίπλοκες αρχιτεκτονικές της βιβλιογραφίας όπως είναι τα Recurrent Neural Networks (RNNs). Η διαφορά αυτή προκύπτει λόγω της ανάγκης για υψηλές επιδόσεις, αφού στα RNNs η ροή των δεδομένων ακολουθεί “κύκλους”, απαιτώντας έτσι πολλές επαναλήψεις που δεν είναι τόσο αποδοτικές. Αντίθετα, στα CNNs η πορεία των δεδομένων είναι γραμμική: από την είσοδο στην έξοδο.

Στην πιο πρόσφατη έκδοση της αρχιτεκτονικής YOLOv3, η οποία χρησιμοποιείται και στα πλαίσια της παρούσας εργασίας, χρησιμοποιούνται συνολικά 53 convolutional layers. Η αρχιτεκτονική αυτή ουσιαστικά προβλέπει τις συντεταγμένες του πλαισίου που δείχνει τη θέση του αντικειμένου στην εικόνα. Έπειτα, για κάθε πλαίσιο, γίνεται πρόβλεψη της κλάσης του, δηλαδή της κατηγορίας του αντικειμένου (π.χ., άνθρωπος, αυτοκίνητο, κλπ), μέσω λογιστικών ταξινομητών. Η ταξινόμηση γίνεται με multilabel τρόπο, δηλαδή ένα πλαίσιο μπορεί να ανήκει σε περισσότερες από μία κατηγορίες, κάτι που κρίνεται πιο αντιπροσωπευτικό από τους ερευνητές. Μέσω αυτής της προσέγγισης, το YOLOv3 έχει παρόμοια απόδοση σε διάφορες μετρικές με άλλα, μεγαλύτερα νευρωνικά δίκτυα, όπως το RetinaNet, έχοντας, όμως, πολύ υψηλότερη ταχύτητα.

Στα πλαίσια της παρούσας εργασίας, το YOLOv3 χρησιμοποιείται σε κάθε ένα frame από τα trailers που έχουν συλλεχθεί. Έπειτα, καταγράφεται αθροιστικά το συνολικό πλήθος εμφανίσεων κάθε αντικειμένου για όλο το trailer. Το πλήθος εμφανίσεων κάθε αντικειμένου και τα αντικείμενα που εμφανίζονται ανά frame αποθηκεύεται σε αρχεία, ώστε να χρησιμοποιηθούν αργότερα για το στάδιο της ταξινόμησης.

Συγκεκριμένα, για να επιτευχθούν τα προαναφερόμενα, αναπτύσσεται πρόγραμμα σε Python 3.7, το οποίο αξιοποιεί τη βιβλιοθήκη ImageAI⁸. Η βιβλιοθήκη αυτή επιτρέπει την ανίχνευση των αντικειμένων σε κάθε frame του βίντεο, χρησιμοποιώντας στο παρασκήνιο την αρχιτεκτονική του YOLOv3, με μία εύχρηστη προγραμματιστική διεπαφή. Επίσης, χρησιμοποιούνται βάρη που έχουν ήδη διαμορφωθεί από διαδικασία εκπαίδευσης στο σύνολο δεδομένων Common Objects in Context (COCO) (Lin et al., 2014), το οποίο απαρτίζεται από 80 κλάσεις αντικειμένων. Ακόμη, αξίζει να αναφερθεί πως, για να ολοκληρωθεί πιο γρήγορα η διαδικασία ανίχνευσης αντικειμένων, τα frames τροφοδοτούνται στο νευρωνικό δίκτυο σε ανάλυση 128x128.

3.1.4 Διαμόρφωση Συνόλου Δεδομένων Ταξινόμησης

Έχοντας ολοκληρώσει τη διαδικασία ανίχνευσης αντικειμένων στις σκηνές ενός trailer, απαιτείται μία διαδικασία προεπεξεργασίας, ώστε να διαμορφωθεί ένα σύνολο δεδομένων που θα χρησιμοποιηθεί στη συνέχεια για τον πειραματισμό με αλγορίθμους ταξινόμησης.

Για να γίνει αυτό, χρησιμοποιούνται τα στοιχεία που έχουν συλλεχθεί για κάθε ταινία στα προηγούμενα βήματα, σε συνδυασμό με το συνολικό πλήθος εμφανίσεων αντικειμένων σε αυτή. Για την ακρίβεια, οι 80 κλάσεις αντικειμένων που μπορεί να προβλέψει η τεχνική YOLO αποτελούν τα χαρακτηριστικά, τα οποία θα χρησιμοποιηθούν ως χαρακτηριστικά εκπαίδευσης για την αναγνώριση συγκεκριμένων προτύπων και αναπαρίστανται ως στήλες στο σύνολο δεδομένων. Αντίστοιχα, κάθε ταινία αποτελεί μία γραμμή, η οποία περιέχει τις 80 τιμές που αντιστοιχούν στις κλάσεις αντικειμένων, την κλάση (ή τις κλάσεις) στην οποία ανήκει, δηλαδή το είδος της ταινίας, καθώς και κάποια βασικά αναγνωριστικά στοιχεία, όπως είναι το αναγνωριστικό ID της ταινίας στο IMDb και του trailer στο YouTube. Ακόμη, καταγράφεται το πλήθος των frames που συνιστούν το trailer, για λόγους που θα γίνουν πιο ξεκάθαροι στην πορεία.

Πίνακας 3.1: Ένα δείγμα του συνόλου δεδομένων, όπως διαμορφώνεται, ώστε να ακολουθήσει η ταξινόμηση των ειδών μίας ταινίας.

| ID | IMDb ID | YouTube ID | Title | Genres | Objects | # of Frames |
|----|-----------|-------------|-----------------|---|---------------------------------|-------------|
| 1 | tt4900708 | iunJULOGRFk | Creepy | ['Crime', 'Drama', 'Horror', 'Mystery', 'Thriller'] | {'person': 130, "tie": 2, ...} | 418 |
| 2 | tt1825636 | Hnm4oedSBk4 | 170 Hz | ['Drama'] | {"person": 452, "cup": 40, ...} | 591 |
| 3 | tt3696192 | J6HIAfNxzQk | Singh Is Bliing | ['Action', 'Comedy'] | {"person": 1194, | 883 |

⁸ <https://imageai.readthedocs.io/en/latest/>

| | | | | | | |
|--|--|--|---------|--|--------------------|--|
| | | | Officia | | "bowl": 2, ...} | |
|--|--|--|---------|--|--------------------|--|

Η δομή του συνόλου δεδομένων παρουσιάζεται συνοπτικά και στον Πίνακα 3.1, όπου εμφανίζονται ενδεικτικά μερικά δείγματα ταινιών. Αξίζει να παρατηρηθεί πως, αν και οι περισσότερες στήλες έχουν μία απλή δομή, οι στήλες Genres (είδη της ταινίας) και Objects (αντικείμενα που ανιχνεύθηκαν) παίρνουν σύνθετες τιμές. Αυτό συμβαίνει, διότι οι στήλες αυτές, λόγω της φύσης τους, έχουν πολλαπλές τιμές, αφού μία ταινία μπορεί να ανήκει σε παραπάνω από ένα είδη, ενώ στα αντικείμενα που ανιχνεύθηκαν πρέπει να καταγραφεί και το όνομα αλλά και το πλήθος εμφανίσεων αυτών. Έτσι, οι στήλες αυτές έχουν ιδιαίτερη μορφοποίηση, η οποία μπορεί εύκολα να μετατραπεί στον αντίστοιχο τύπο δεδομένων μέσω Python. Δηλαδή, η στήλη Genres αποθηκεύεται με τη μορφή που αναπαρίσταται μία λίστα στην Python, ενώ η στήλη Objects όπως ένα αρχείο JSON.

3.2 Ταξινόμηση του Είδους των Ταινιών

Οι αλγόριθμοι ταξινόμησης είναι επίσης ένας κλάδος της TN που έχει γνωρίσει μεγάλη άνθηση τις τελευταίες δεκαετίες, χάρη στη χρήση των DNNs. Άλλωστε, ακόμα και η αναγνώριση αντικειμένων που αναλύθηκε προηγουμένως απαιτεί τη χρήση αλγορίθμων ταξινόμησης για την εξακρίβωση του πλαισίου και της κλάσης κάθε αντικειμένου. Σε αυτό το χρονικό διάστημα, διάφορες νέες τεχνικές έχουν αναπτυχθεί, αλλά και παλαιότερες έχουν βελτιωθεί, οι οποίες εστιάζουν τόσο στην βελτιστοποίηση των ίδιων των αλγορίθμων, αλλά και στη βέλτιστη προεπεξεργασία των δεδομένων, με στόχο πάντα την αποτελεσματικότερη ταξινόμηση.

Παρόλο που ως κλάδος έχει αναπτυχθεί αρκετά χάρη στα DNNs, ενδιαφέρον παρουσιάζει το γεγονός ότι αυτά προσφέρουν κυρίως καλύτερες και αυτοματοποιημένες τεχνικές αναπαράστασης των δεδομένων, ή εξαγωγής χαρακτηριστικών, ουσιαστικά. Για τον λόγο αυτόν, τα περισσότερα DNNs δέχονται ως είσοδο ακατέργαστα δεδομένα, όπως είναι οι εικόνες, αντί για προϋπολογισμένα χαρακτηριστικά, και επιστρέφουν ως έξοδο την εκάστοτε κλάση. Αντίθετα, αποκλειστικά και μόνο για τους σκοπούς της ταξινόμησης, χρησιμοποιούνται συνήθως πλήρως συνδεδεμένα layers νευρώνων, με κατάλληλες συναρτήσεις σφάλματος, ανάλογα με το είδος της ταξινόμησης (Miikkulainen et al., 2019).

Έτσι, η επιλογή μεταξύ κλασικών αλγορίθμων και πιο σύγχρονων, όπως τα DNNs, συχνά έχει να κάνει με το είδος της εισόδου που θα τροφοδοτηθεί στον αλγόριθμο ταξινόμησης και το πόσο αποδοτική είναι η αναπαράσταση αυτής, ώστε να γίνει η ταξινόμηση. Ακόμη, σημαντικό μειονέκτημα στη χρήση συγκεκριμένων αλγορίθμων, όπως τα DNNs, είναι πως δεν είναι εύκολη η ερμηνεία του γιατί επιλέγεται μία κλάση αντί μίας άλλης. Εξίσου καθοριστικό ρόλο παίζει συχνά και η διαθέσιμη υπολογιστική ισχύς, αφού τα DNNs απαιτούν συνήθως κάρτες γραφικών υψηλών επιδόσεων, με αντίστοιχο πάντα χρηματικό κόστος (Ongsulee, 2017).

Συνεπώς, απαραίτητο είναι να επιλεγούν οι κατάλληλοι αλγόριθμοι ταξινόμησης, έχοντας, πάντα, υπόψη τους στόχους της παρούσας εργασίας. Μιας και το κύριο ερώτημα προς

απάντηση είναι το κατά πόσο μπορούν να χρησιμοποιηθούν τα αντικείμενα που εμφανίζονται στις σκηνές μιας ταινίας, ώστε να προβλεφθεί το είδος αυτής, τα δεδομένα που θα χρησιμοποιηθούν ως είσοδος για τον όποιο αλγόριθμο έχουν ήδη αποφασιστεί. Η είσοδος είναι, δηλαδή, το πλήθος των εμφανίσεων κάθε αντικειμένου στις σκηνές (ή τα frames) της ταινίας. Με την απαραίτητη προεπεξεργασία (η οποία περιγράφεται στη συνέχεια), τα δεδομένα εισόδου θα βρίσκονται σε πινακοειδή (tabular) μορφή, όπου οι κλασικοί αλγόριθμοι ή τα απλά νευρωνικά, όπως είναι τα Multilayer Perceptrons (MLPs), μπορούν να αποδώσουν ικανοποιητικά και δε χρειάζονται πιο περίπλοκες αρχιτεκτονικές, γι' αυτό και προτιμούνται στα πλαίσια της παρούσας εργασίας.

Παρ' όλ' αυτά, ένα ιδιαίτερο χαρακτηριστικό που παρουσιάζει η ταξινόμηση ταινιών είναι πως μία ταινία ανήκει πιθανώς σε περισσότερα από ένα είδη, γεγονός που περιπλέκει την κατηγοριοποίησή της. Για την αντιμετώπισή του αυτού του προβλήματος με πιο απλοϊκό τρόπο, διάφορες προσεγγίσεις έχουν δοκιμαστεί στη βιβλιογραφία (Zhou et al., 2010, Simões et al., 2016). Όμως, η απλούστευση του προβλήματος, αν και παρέχει καλύτερη απόδοση βάσει μετρικών, έχει λάβει κριτική ότι δεν ανταποκρίνεται στην πραγματικότητα. Για παράδειγμα, η παραδοχή της δημοσίευσης (Zhou et al., 2010) ότι αρκεί ένας αλγόριθμος να πετύχει μόνο ένα είδος μιας ταινίας, ώστε να θεωρηθεί σωστή η πρόβλεψη, έχει ως αποτέλεσμα οι προβλέψεις σε ταινίες που ανήκουν σε 3 είδη από τα 4 είδη ταινιών που εξετάζουν να είναι σωστές με 75% πιθανότητα, κάτι που ισχύει για το 54% του συνόλου δεδομένων που χρησιμοποιούν.

Για να αποφευχθούν τέτοιου είδους σφάλματα, η προσέγγιση που ακολουθείται σε αυτή την εργασία δε χρησιμοποιεί καμία παραδοχή για την απλούστευση του προβλήματος, όπως γίνεται και στην έρευνα (Wehrmann & Barros, 2017). Συγκεκριμένα, μελετώνται συνολικά 5 είδη ταινιών: Drama, Comedy, Romance, Action, Thriller. Μία ταινία μπορεί να ανήκει σε παραπάνω από ένα είδος και, όπως προκύπτει από τα δεδομένα που συλλέγονται και έχουν περιγραφεί παραπάνω, μπορεί να ανήκει το πολύ σε 3 είδη. Για το λόγο αυτό, η ταξινόμηση θεωρείται multilabel, με κάθε είδος να θεωρείται μία ξεχωριστή κλάση, η οποία παίρνει τιμές 0 ή 1. Τα multilabel προβλήματα ταξινόμησης θεωρούνται γενικά δυσκολότερα από τα απλούστερα προβλήματα που αποτελούνται από μία κλάση, και η αξιολόγησή τους γίνεται πιο περίπλοκη.

3.2.1 Multi-label Ταξινόμηση

Τα κοινότυπα προβλήματα ταξινόμησης αφορούν κατά κόρον δείγματα που ανήκουν σε μία μόνο κλάση. Η κλάση αυτή μπορεί να δέχεται δύο μόνο τιμές, οπότε το πρόβλημα ταξινόμησης χαρακτηρίζεται ως δυαδικό (binary), ή παραπάνω από δύο, οπότε το πρόβλημα χαρακτηρίζεται ως multiclass. Για παράδειγμα, στην περίπτωση που γίνεται ταξινόμηση για το αν μία ταινία ανήκει στο είδος Thriller, η απάντηση θα είναι ναι ή όχι, οπότε αποτελεί binary ταξινόμηση. Αντίστοιχα, αν μία ταινία χαρακτηρίζεται μόνο από ένα είδος και δεν μπορεί να ανήκει ταυτόχρονα σε παραπάνω από ένα, η κλάση θα έπαιρνε μία από τις τιμές Drama, Comedy, Romance, Action, Thriller και η ταξινόμηση θα ήταν multiclass.

Όμως, στην πραγματικότητα, ένα σημαντικό ποσοστό των ταινιών ανήκουν σε παραπάνω από ένα είδη ταυτόχρονα. Έτσι, ο καλύτερος τρόπος αναπαράστασης είναι ένα δείγμα να χαρακτηρίζεται από πολλές κλάσεις, που παίρνουν binary τιμές, πρόβλημα που χαρακτηρίζεται ως multi-label ταξινόμηση, το οποίο αποτελεί γενίκευση της multiclass ταξινόμησης. Η multi-label ταξινόμηση μπορεί να εκφραστεί μαθηματικά, όσο πιο απλά γίνεται με τον εξής τρόπο: δεδομένων των εισόδων x αναζητείται ένα μοντέλο που να τις αντιστοιχεί σε δυαδικά διανύσματα y , αναθέτοντας μία τιμή 0 ή 1 για κάθε κλάση στο y .

Ο μάλλον πιο απλός τρόπος αντιμετώπισης των multi-label προβλημάτων ταξινόμησης είναι η μετατροπή του προβλήματος σε binary ταξινόμηση. Δηλαδή, διασπάται το multi-label πρόβλημα σε πολλά binary προβλήματα. Με τον τρόπο αυτό, χρειάζεται να εκπαιδευτεί ένας ταξινομητής ανά κλάση ανεξάρτητα από τις υπόλοιπες. Έτσι, η τελική πρόβλεψη αποτελείται από τις επιμέρους προβλέψεις των ταξινομητών για κάθε κλάση. Το κύριο μειονέκτημα αυτής της μεθόδου προκύπτει επειδή ακριβώς η εκπαίδευση κάθε ταξινομητή γίνεται ανεξάρτητα από τις υπόλοιπες κλάσεις. Συνεπώς, δε λαμβάνονται υπόψη ιδιαίτερα χαρακτηριστικά, όπως συσχετίσεις μεταξύ των διαφορετικών κλάσεων. Ένας τρόπος αντιμετώπισης αυτού του προβλήματος είναι η διαμόρφωση μίας αλληλουχίας από ταξινομητές, όπου ο καθένας δέχεται ως είσοδο την πρόβλεψη του προηγούμενου, όμως αυτή η μέθοδος δεν εξετάζεται εδώ.

Μέσω της διάσπασης του multi-label προβλήματος σε πολλαπλά προβλήματα binary ταξινόμησης, μπορούν να αξιοποιηθούν οι συνήθεις τεχνικές που χρησιμοποιούνται για τα τελευταία. Εξίσου, μπορούν να χρησιμοποιηθούν οι κλασικές μετρικές αξιολόγησης των binary ταξινομητών για κάθε κλάση, ενώ διάφορες, πιο περίπλοκες εναλλακτικές έχουν προταθεί για τη συνολική αξιολόγηση του σύνθετου ταξινομητή, οι οποίες αναλύονται περαιτέρω στο Κεφάλαιο 4.1.

Στη συνέχεια, αναλύονται οι διάφορες τεχνικές που εφαρμόζονται, συμπεριλαμβανομένης της προεπεξεργασίας του συνόλου δεδομένων, ώστε να γίνει πειραματισμός με αλγόριθμους ταξινόμησης.

3.2.2 Προεπεξεργασία

Το σύνολο δεδομένων, όπως αυτό έχει περιγραφεί προηγουμένως, περιέχει ταινίες χωρίς κάποια διάκριση ως προς το είδος τους. Επομένως, σε αυτή την ενότητα δικαιολογείται το πώς και γιατί επιλέγονται τα 5 είδη ταινιών που χρησιμοποιούνται τελικά. Ακόμη, οι στήλες Objects και Genres του Πίνακα 3.1 περιέχουν σχετικά “αδόμητα” δεδομένα. Με στόχο τον πειραματισμό με διαφορετικούς ταξινομητές, είναι απαραίτητη η προεπεξεργασία αυτών των δεδομένων, ώστε να φτάσουν τελικά στην κατάλληλη πινακοειδή μορφή.

Για να επιλεγεί, τελικά, το σύνολο των ειδών που θα διατηρηθεί στο σύνολο δεδομένων προς ταξινόμηση, πρέπει να γίνουν ξεκάθαρα τα κριτήρια επιλογής αυτών. Αρχικά, λόγω του γεγονότος ότι πρόκειται για multi-label πρόβλημα ταξινόμησης, όσο περισσότερα είδη ταινιών υπάρχουν, τόσο πιο δύσκολη θα είναι η εκπαίδευση ενός αποτελεσματικού ταξινομητή. Αυτό συμβαίνει, εν μέρει, διότι απαιτείται ένα ικανοποιητικό σύνολο δεδομένων για κάθε κλάση.

Για παράδειγμα, αν γίνει η υπόθεση ότι χρειάζονται ενδεικτικά 100 ταινίες που να ανήκουν σε ένα είδος, ώστε ο ταξινομητής να έχει ικανοποιητικά αποτελέσματα, τότε θα χρειαστούν 1000 ταινίες για 10 διαφορετικά είδη ταινιών. Αν και η υπόθεση αυτή δεν ανταποκρίνεται απόλυτα στην πραγματικότητα, είναι ξεκάθαρο πως, ανάλογα με το πλήθος των ειδών που θα επιλεγούν, θα αυξάνεται γραμμικά και το απαιτούμενο πλήθος των δεδομένων.

Το γεγονός ότι το μέγεθος του συνόλου δεδομένων πρέπει να αυξάνεται ανάλογα με το πλήθος των ειδών προκαλεί και ένα άλλο ζήτημα. Με βάση το προηγούμενο παράδειγμα, καθώς εισάγονται περισσότερες κλάσεις, τόσο πιο ανομοιόμορφα κατανομημένα (imbalanced) θα είναι τα δείγματα στις κλάσεις για τον κάθε ταξινομητή. Δηλαδή, με βάση το προηγούμενο παράδειγμα, όπου υπάρχουν 10 διαφορετικά είδη ταινιών με 100 ταινίες στο καθένα, για έναν ταξινομητή που λαμβάνει υπόψη μόνο το ένα είδος, θα υπάρχουν 100 δείγματα που ανήκουν στο είδος και 900 που δεν θα υπάρχουν. Έτσι, είναι αρκετά εύκολο οι προβλέψεις του ταξινομητή να είναι “ευνοϊκές” προς την αρνητική πρόβλεψη, ότι δηλαδή μία ταινία δεν ανήκει στο είδος, για το οποίο είναι υπεύθυνος να προβλέψει ο ίδιος ο ταξινομητής. Άλλωστε, αν όλες οι προβλέψεις ήταν αρνητικές, ο ταξινομητής θα είχε ποσοστά επιτυχίας (accuracy) της τάξης του 90%.

Το πρόβλημα αυτό του imbalancing δεν είναι εύκολο να αντιμετωπιστεί με συνηθισμένες τεχνικές όπως το undersampling ή το oversampling, διότι αν αφαιρεθούν ή προστεθούν επιπλέον δείγματα, δεν θα επηρεαστεί μόνο μία κλάση, δεδομένου πάντα ότι μία ταινία δεν ανήκει μόνο σε ένα είδος. Έτσι, προσπαθώντας να προσθέσει κανείς, για παράδειγμα, δείγματα στην κλάση Thriller, θα αύξανε με μεγάλη πιθανότητα ταυτόχρονα και την κλάση Horror. Για τον λόγο αυτό, το πρόβλημα του imbalancing θα πρέπει να αντιμετωπιστεί από τον αλγόριθμο ταξινόμησης κι όχι μέσω επεξεργασίας του συνόλου δεδομένων.

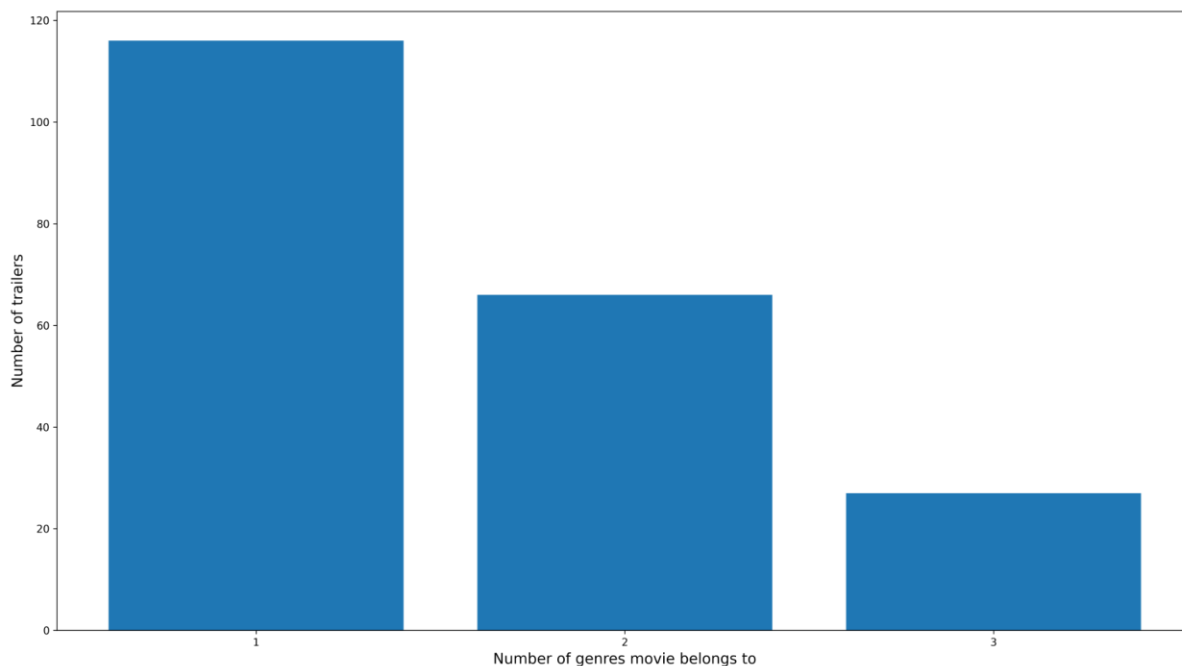
Για τους προαναφερόμενους λόγους, και λαμβάνοντας υπόψη τη βιβλιογραφία, το σύνολο των κλάσεων που επιλέγεται αποτελείται από 5 κλάσεις και είναι οι εξής: Drama, Comedy, Romance, Action, Thriller. Τα είδη αυτά, όπως φαίνεται και στο Σχήμα 3.3, συνιστούν τα 5 πιο συχνά είδη ταινιών στο σύνολο δεδομένων που συλλέχθηκε και συνολικά παρουσιάζει μεγάλη ομοιότητα με τις περισσότερες μελέτες της βιβλιογραφίας. Για παράδειγμα, η έρευνα των Huang & Wang εστιάζει στα ακόλουθα είδη: Drama, Comedy, Action, Horror και Thriller, παρουσιάζοντας μεγάλη ομοιότητα και με αυτή των Simões et al. που επικεντρώνεται σε ταινίες Action, Horror, Comedy και Drama (Huang & Wang, 2012, Simões et al., 2016). Επιπλέον, πρόσθετες έρευνες έχουν ως επίκεντρο τις ταινίες Comedy, Action, Horror, Sci-fi, Crime και Romance (Shambharkar & Doja, 2020, Shambharkar et al., 2020).

Διατηρώντας ταινίες που ανήκουν σε τουλάχιστον 1 από τα 5 επιλεγμένα είδη, στο σύνολο δεδομένων παραμένουν συνολικά 209 από τις 254 ταινίες. Μη λαμβάνοντας υπόψη, δηλαδή, 18 είδη ταινιών, αφαιρούνται μόνο 45 ταινίες από το σύνολο δεδομένων, οπότε είναι ξεκάθαρο πως υπάρχει μεγάλη προτίμηση στα 5 είδη ταινιών που επιλέχθηκαν. Αυτό οφείλεται, εν μέρει, στο γεγονός ότι, π.χ., μία Sci-Fi ή Adventure ταινία, θα ανήκει με μεγάλη πιθανότητα και στο είδος Action. Έτσι, ακόμη κι αν αγνοηθούν τα είδη Sci-Fi και Adventure, οι περισσότερες ταινίες παραμένουν στο σύνολο δεδομένων και κατανέμονται κυρίως στο είδος Action.

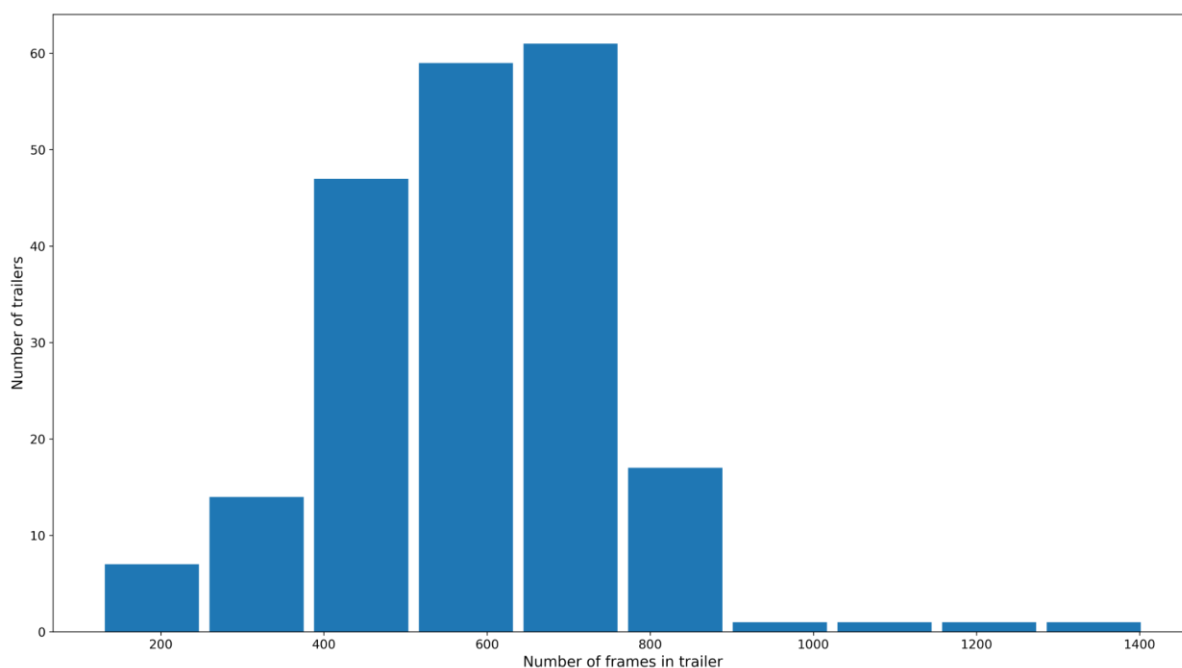
Έχοντας επιλέξει τα είδη των ταινιών και δεδομένου του προαναφερόμενου φαινομένου που προκύπτει αν αγνοηθούν τα περισσότερα είδη ταινιών, είναι ενδιαφέρον να παρατηρηθεί πως μεταβάλλονται κάποια χαρακτηριστικά του συνόλου δεδομένων. Αρχικά, όπως φαίνεται στο Σχήμα 3.6, όλες οι ταινίες ανήκουν πλέον το πολύ σε 3 είδη. Επίσης, συγκριτικά με το Σχήμα 3.4, οι περισσότερες ταινίες ανήκουν πλέον σε μόνο 1 είδος, το οποίο είναι λογικό, αν λάβει κανείς υπόψη το παράδειγμα με τα είδη Sci-Fi και Adventure. Συνοπτικά, τα μεγέθη αυτού του Σχήματος, καθώς και το πλήθος των ταινιών ανά κατηγορία, καταγράφονται στον Πίνακα 3.2 για λόγους ακριβείας. Ακόμη, στο Σχήμα 3.7 παρουσιάζεται και πάλι το ιστόγραμμα του πλήθους των frames στα trailers των ταινιών που παραμένουν στο σύνολο δεδομένων, μετά την αφαίρεση των υπόλοιπων ειδών. Όπως ίσχυε και πριν από αυτή (Σχήμα 3.5), τα περισσότερα trailers αποτελούνται από περίπου 600 frames, οπότε και η κατανομή δεν έχει επηρεαστεί ιδιαίτερα από την αφαίρεση των 45 ταινιών.

Πίνακας 3.2: Κατηγορίες ταινιών στο αρχικό σύνολο δεδομένων που συλλέγεται, προτού υποστεί επεξεργασία.

| Κατηγορίες Ταινιών | Πλήθος |
|--------------------|--------|
| Thriller | 56 |
| Comedy | 66 |
| Drama | 119 |
| Romance | 46 |
| Action | 62 |
| 1 είδους | 116 |
| 2 ειδών | 66 |
| 3 ειδών | 27 |



Σχήμα 3.6: Το πλήθος των ταινιών συναρτίζει του πλήθος των κατηγοριών στο οποίο ανήκουν τελικό σύνολο δεδομένων προς ταξινόμηση.



Σχήμα 3.7: Ιστόγραμμα του πλήθους των frames στις ταινίες που απαρτίζουν το τελικό σύνολο δεδομένων προς ταξινόμηση.

Μιας και το σύνολο των ειδών που συμπεριλαμβάνονται στο σύνολο δεδομένων έχει επιλεγεί, απαιτείται ένα ακόμη βήμα προεπεξεργασίας, προτού αυτό να μπορεί να χρησιμοποιηθεί για την εκπαίδευση και αξιολόγηση αλγορίθμων ταξινόμησης. Συγκεκριμένα, απαιτείται η μετατροπή του συνόλου δεδομένων σε πινακοειδή μορφή, αφού οι στήλες Objects και Genres δεν έχουν μόνο μία τιμή ανά δείγμα, αλλά διατηρούν την πληροφορία είτε σε μορφή λίστας

είτε σε JSON. Και για τις δύο στήλες, η μέθοδος προεπεξεργασίας είναι αρκετά όμοια κι έχει ως στόχο τη μετατροπή των πιθανών τιμών σε πολλαπλές στήλες, με μία στήλη για κάθε τιμή.

Κάτι που αξίζει να αναφερθεί σε αυτό το σημείο είναι πως τα χαρακτηριστικά, όπως διαμορφώνονται από το προηγούμενο στάδιο προεπεξεργασίας, αποτελούν το γνωστό μοντέλο αναπαράστασης πληροφορίας Bag of Words. Το μοντέλο αυτό προέρχεται από το χώρο της Γλωσσικής Επεξεργασίας και χρησιμοποιείται για την αναπαράσταση ενός κειμένου μέσω του πλήθους εμφανίσεων των λέξεων που το συνιστούν. Το εν λόγω μοντέλο έχει εφαρμοστεί και σε άλλους τομείς, όπως είναι η Μηχανική Όραση. Όμως, ένα μειονέκτημα του μοντέλου αυτού, ανεξάρτητα από τον τομέα εφαρμογής του, είναι πως αυξάνεται σημαντικά η διαστασιμότητα του προβλήματος. Ακόμη, το μοντέλο δεν αναπαριστά με κάποιον τρόπο τις συσχετίσεις μεταξύ των αντικειμένων που αναπαριστά, όπως είναι η σειρά των λέξεων σε μία πρόταση ή η χωρική εγγύτητα δύο αντικειμένων σε μία εικόνα.

Συγκεκριμένα, η στήλη Objects περιέχει το συνολικό πλήθος εμφανίσεων κάθε αντικειμένου, όπως προκύπτει από την τεχνική YOLO. Έτσι, μπορεί να περιέχει μέχρι και 80 διαφορετικά αντικείμενα για κάθε ταινία, αν και συνήθως είναι αρκετά λιγότερα αυτά που ανιχνεύονται. Έτσι, κάθε αντικείμενο μετατρέπεται σε μία ξεχωριστή στήλη, με το συνολικό πλήθος εμφανίσεων αυτού να εξάγεται από τη μορφή JSON και να εισάγεται στην κατάλληλη στήλη. Αντίστοιχα, για τη στήλη Genres, δημιουργούνται 5 νέες στήλες για καθένα από τα είδη που επιλέγονται και εισάγεται η τιμή 1 στην εκάστοτε στήλη, αν η ταινία ανήκει στο αντίστοιχο είδος. Με τον τρόπο αυτόν, το σύνολο δεδομένων αποκτά τη μορφή που φαίνεται στον Πίνακα 3.3. Αξίζει να αναφερθεί πως οι στήλες ID και IMDb ID δε χρησιμοποιούνται ως χαρακτηριστικά στην ταξινόμηση, αφού αποτελούν μοναδικά αναγνωριστικά.

3.2.3 Αλγόριθμοι

Έχοντας ολοκληρώσει τη διαδικασία προεπεξεργασίας του συνόλου δεδομένων, είναι δυνατό να εφαρμοστούν αλγόριθμοι ταξινόμησης, ώστε να προβλέπουν το είδος μιας ταινίας με βάση το συνολικό πλήθος εμφανίσεων των διαφόρων αντικειμένων που ανιχνεύονται στις σκηνές του trailer αυτής. Έτσι, στη συνέχεια αναλύονται οι διαφορετικοί αλγόριθμοι ταξινόμησης και οι παράμετροι αυτών που χρησιμοποιούνται για την διεξαγωγή πειραμάτων, καθώς και άλλα προπαρασκευαστικά βήματα που πιθανώς χρειάζονται για την εφαρμογή τους.

Πίνακας 3.3: Ένα δείγμα του τελικού συνόλου δεδομένων, όπως διαμορφώνεται, ώστε να ακολουθήσει η ταξινόμηση των ειδών ταινιών.

| ID | IMDb ID | # of Frames | ... | person | ... | Thriller | Comedy | Drama | Romance | Action |
|----|-----------|-------------|-----|--------|-----|----------|--------|-------|---------|--------|
| 1 | tt4900708 | 418 | | 130 | | 1 | 0 | 1 | 0 | 0 |
| 2 | tt1825636 | 591 | | 452 | | 0 | 0 | 1 | 0 | 0 |
| 3 | tt | 883 | | 1194 | | 0 | 1 | 0 | 0 | 1 |

| | | | | | | | | | |
|--|---------|--|--|--|--|--|--|--|--|
| | 3696192 | | | | | | | | |
|--|---------|--|--|--|--|--|--|--|--|

Το πρώτο και κυριότερο προπαρασκευαστικό βήμα, το οποίο αφορά το σύνολο δεδομένων και απαιτείται για όλους τους αλγορίθμους ταξινόμησης που ακολουθούν, είναι η κανονικοποίηση. Συγκεκριμένα, αξίζει να παρατηρήσει κανείς πως αν ένα trailer είναι πολύ μεγαλύτερο σε μήκος από ένα άλλο, είναι σχεδόν βέβαιο πως συνολικά θα ανιχνευθούν και περισσότερα αντικείμενα. Με άλλα λόγια, σχεδόν όλα τα χαρακτηριστικά μάθησης παρουσιάζουν υψηλή συσχέτιση με το μήκος του βίντεο. Αυτό δεν ωφελεί ιδιαίτερα τον όποιο αλγόριθμο ταξινόμησης, καθώς υπάρχει ο κίνδυνος να συσχετιστεί το μήκος του βίντεο με το είδος της ταινίας, κάτι που δε συμβαδίζει με το σκοπό της εργασίας.

Για να απαλειφθεί αυτή η συσχέτιση, πραγματοποιείται κανονικοποίηση όλων των αντικειμένων που ανιχνεύονται και αποτελούν χαρακτηριστικά. Για την ακρίβεια, το συνολικό πλήθος εμφανίσεων κάθε αντικειμένου διαιρείται με το πλήθος των frames. Με τον τρόπο αυτό, τα χαρακτηριστικά δεν υποδεικνύουν ακριβώς το συνολικό πλήθος εμφανίσεων, αλλά το πλήθος εμφανίσεων ανά frame. Να σημειωθεί πως η κανονικοποίηση αυτή δεν έχει ως αποτέλεσμα τα χαρακτηριστικά να ανήκουν στο εύρος $[0, 1]$, διότι ένα αντικείμενο μπορεί να εμφανίζεται πολλές φορές σε ένα frame.

Συνεχίζοντας με τους αλγορίθμους ταξινόμησης, επιλέγεται μία πληθώρα αυτών για πειραματισμό για τους οποίους ακολουθεί μία σύντομη περιγραφή και κάποιοι ενδεικτικοί λόγοι για τους οποίους επιλέχθηκαν. Οι αλγόριθμοι αυτοί έχουν επιλεγθεί σύμφωνα με τη βιβλιογραφία⁹ (Fei & Zhang, 2019, Alrowaili, 2019). Αξίζει να σημειωθεί πως η εκπαίδευση και η αξιολόγηση της ταξινόμησης γίνεται μέσω προγράμματος σε Python 3.7 με τη βοήθεια της βιβλιοθήκης Scikit-learn (Pedregosa et al., 2011). Η συγκεκριμένη βιβλιοθήκη παρέχει υλοποιήσεις μίας μεγάλης γκάμας “κλασικών” αλγορίθμων Μηχανικής Μάθησης, ενώ δεν εστιάζει σε DNNs.

Αρχικά, για την αξιολόγηση των αλγορίθμων, είναι μία συνετή ιδέα να υπάρχει ένας απλός αλγόριθμος που θα λειτουργεί ως μέτρο σύγκρισης για όλους τους υπόλοιπους. Για το σκοπό αυτό, χρησιμοποιείται ο αλγόριθμος Dummy Classifier του scikit-learn, ο οποίος παράγει προβλέψεις με απλούς κανόνες, αντί να εκπαιδεύεται όντως κάποιο σύνθετο μοντέλο με βάση τα χαρακτηριστικά μάθησης.

Για την ακρίβεια, με το συγκεκριμένο αλγόριθμο γίνεται πειραματισμός χρησιμοποιώντας τις εξής παραμέτρους για τους κανόνες πρόβλεψης:

- **“stratified”**, γίνεται πρόβλεψη με βάση την κατανομή των δειγμάτων στις κλάσεις
- **“most_frequent”**, γίνεται πάντα πρόβλεψη της πιο συχνής κλάσης
- **“uniform”**, παράγει προβλέψεις τυχαία ανάμεσα στις τιμές της κλάσεις (κάθε κλάση είναι ισοπίθανη)

⁹ http://vision.stanford.edu/teaching/cs131_fall1718/files/14_BoW_bayes.pdf

Ένας ακόμη αλγόριθμος ταξινόμησης που χρησιμοποιείται είναι ο Naive Bayes. Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί δεσμευμένες πιθανότητες για να υπολογίσει την πιθανότητα ένα δείγμα να ανήκει σε μία κλάση, δεδομένων των χαρακτηριστικών που το περιγράφουν. Είναι ένας αρκετά απλός αλγόριθμος, γι' αυτό και συχνά δεν είναι ο βέλτιστος τρόπος επίλυσης ενός προβλήματος, αλλά αποδίδει ικανοποιητικά καλά σε πολλές περιπτώσεις. Ένα μειονέκτημα του “παραδοσιακού” Naive Bayes είναι πως δεν μπορεί να αντιμετωπίσει εύκολα συνεχείς τιμές στα χαρακτηριστικά μάθησης. Για την αντιμετώπιση αυτού, εδώ χρησιμοποιείται μία επέκτασή του αλγορίθμου, γνωστή ως Gaussian Naive Bayes, η οποία υποθέτει πως τα χαρακτηριστικά ακολουθούν κανονική κατανομή στον υπολογισμό των πιθανοτήτων.

Ένας ακόμη αλγόριθμος που χρησιμοποιείται είναι τα Δέντρα Απόφασης (Decision Trees). Τα Decision Trees είναι μία ιδιαίτερα διαδεδομένη κατηγορία αλγορίθμων ταξινόμησης, χάρη στην ιδιότητα ότι το μοντέλο που διαμορφώνουν δε χρησιμοποιεί αριθμητικές παραμέτρους, αλλά μία σειρά από λογικούς κανόνες που συνιστούν δέντρα αποφάσεων. Λόγω αυτού, τα αποτελέσματά τους είναι εύκολα ερμηνεύσιμα, γεγονός που συμβάλλει στην αξιοπιστία των μοντέλων, αφού μπορεί να ελεγχθεί εύκολα αν κάποιος από τους λογικούς κανόνες δε συμβαδίζει με την πραγματικότητα.

Στα Decision Trees υπάρχει μία μεγάλη γκάμα παραμέτρων που επηρεάζουν τον τρόπο κατασκευής τους, οι οποίες χρησιμοποιούνται εδώ για πειραματισμό:

- **criterion**, το κριτήριο με βάση το οποίο γίνεται ο διαχωρισμός των δεδομένων και κατ' επέκταση οι διακλαδώσεις του δέντρου. Παίρνει τιμές “gini” και “entropy”.
- **max_depth**, το μέγιστο βάθος του δέντρου. Χρησιμοποιείται κυρίως για να απλοποιηθεί το δέντρο, ώστε να είναι πιο εύκολα ερμηνεύσιμο και να αποφευχθεί το overfitting.
- **class_weight**, ανάθεση βαρών σε κάθε κλάση του προβλήματος. Με την προεπιλεγμένη τιμή του αλγορίθμου (None), χρησιμοποιείται ως βάρος η τιμή 1 για κάθε κλάση. Εναλλακτικά, χρησιμοποιείται η τιμή “balanced”, η οποία αναθέτει βάρη σε κάθε κλάση, ανάλογα με την κατανομή των δειγμάτων στις κλάσεις, όπως προκύπτει από τα δεδομένα εκπαίδευσης.

Επίσης, χρησιμοποιείται ο αλγόριθμος Random Forest, ο οποίος δημιουργεί ένα πλήθος από Decision Trees, το καθένα από τα οποία εκπαιδεύεται σε ένα υποσύνολο των δεδομένων αλλά και των χαρακτηριστικών εκπαίδευσης. Συνεπώς, η τελική πρόβλεψη προκύπτει ως ο μέσος όρος των προβλέψεων των επιμέρους δέντρων. Με τον τρόπο αυτόν, αντιμετωπίζεται ένα από τα σημαντικότερα μειονεκτήματα των Decision Trees, ότι μπορεί να παραχθούν πολύ διαφορετικά δέντρα από τη διαδικασία εκπαίδευσης αν γίνουν ακόμη και μικρές αλλαγές στα δεδομένα.

Ο αλγόριθμος Random Forest, μιας και βασίζεται σε Decision Trees, έχει αρκετές κοινές παραμέτρους με αυτά, αλλά και κάποιες διαφορετικές, οι οποίες χρησιμοποιούνται στον πειραματισμό και είναι:

- **criterion** και **max_depth**, όπως και στα Decision Trees.

- **class_weight**, επηρεάζει κι εδώ τα βάρη κάθε κλάσης. Σε αντίθεση με τα Decision Trees, εδώ υπάρχει ακόμα η επιλογή “balanced_subsample” βάσει της οποίας, σε κάθε κλάση ανατίθεται ένα βάρος ανάλογο με την κατανομή των δειγμάτων στις κλάσεις, όπως προκύπτει σε κάθε υποσύνολο που τροφοδοτείται σε κάθε δέντρο, αντί για ολόκληρο το σύνολο δεδομένων εκπαίδευσης.
- **n_estimators**, το πλήθος των δέντρων που χρησιμοποιούνται για ταξινόμηση. Καθώς αυξάνεται αυτή η παράμετρος, επηρεάζεται αρνητικά η απόδοση του αλγορίθμου ως προς το χρόνο, αλλά (αν και απλός κανόνας) τόσο πιο αντιπροσωπευτική η αποτελεσματικότητά της ταξινόμησης.

Μία λίγο πιο περίπλοκη κατηγορία αλγορίθμων που χρησιμοποιείται κατά τον πειραματισμό είναι τα SVMs. Ένα από τα πλεονεκτήματα των SVMs είναι πως παρέχουν καλή απόδοση σε σύνολα δεδομένων με πολλά χαρακτηριστικά μάθησης, όπως εδώ, ακόμη κι αν το πλήθος των δειγμάτων δεν είναι ιδιαίτερα μεγάλο σε σχέση με το πλήθος των χαρακτηριστικών. Ως προς την υλοποίηση, εδώ χρησιμοποιείται ο αλγόριθμος Support Vector Classifier (SVC) της βιβλιοθήκης Scikit-learn.

Τα SVMs, όμως, αξιοποιούν μία γκάμα παραμέτρων, η οποία μπορεί να επηρεάσει σημαντικά την απόδοση του αλγορίθμου και οι τιμές της συχνά πρέπει να οριστούν πειραματικά. Οι παράμετροι που λαμβάνονται υπόψη κατά τον πειραματισμό είναι οι ακόλουθες:

- **kernel**, καθορίζει το kernel που θα χρησιμοποιεί ο αλγόριθμος. Χρησιμοποιείται η προεπιλεγμένη τιμή “rbf”, η οποία αντιστοιχεί στο kernel Radial Basis Function.
- **C**, παράμετρος που καθορίζει το regularization που εφαρμόζεται από τον αλγόριθμο. Η τιμή της καθορίζεται πειραματικά βάσει της απόδοσης του αλγορίθμου.
- **class_weight**, καθορίζει το βάρος που ανατίθεται σε κάθε κλάση. Ομοίως με τα Decision Trees και το Random Forest, μέσω της επιλογής “balanced”, το βάρος κάθε κλάσης είναι ανάλογο της κατανομής των τιμών στις κλάσεις, όπως προκύπτει από το σύνολο δεδομένων.

Τέλος, γίνεται πειραματισμός με ρηγά νευρωνικά δίκτυα, τύπου Multi-layer Perceptron (MLP). Ένα δίκτυο MLP συνίσταται από μία ακολουθία επιπέδων που αποτελούνται από νευρώνες, οι οποίοι είτε ενεργοποιούνται είτε όχι, ανάλογα με το βάρος και την πόλωση του καθενός. Έτσι, ανάλογα με τα δεδομένα εισόδου, μεταφέρεται ένα σήμα μέχρι το τελευταίο επίπεδο, το οποίο είναι ειδικά σχεδιασμένο, ώστε να προβλέπει την κλάση. Ένα σημαντικό πλεονέκτημα των νευρωνικών δικτύων σε αυτή την περίπτωση είναι πως μπορεί να σχεδιαστεί με κατάλληλο τρόπο το επίπεδο εξόδου, ώστε να παράγουν multi-label προβλέψεις, χωρίς τη χρήση τεχνικών, όπως η μετατροπή σε binary ταξινόμηση.

Έτσι, η αρχιτεκτονική του δικτύου αποτελείται σίγουρα από δύο επίπεδα, το επίπεδο εισόδου και το επίπεδο εξόδου, καθώς και μία ακολουθία από κρυφά επίπεδα, όπως απεικονίζεται στο Σχήμα 3.8. Το επίπεδο εισόδου αποτελείται από ένα συγκεκριμένο πλήθος νευρώνων, το οποίο καθορίζεται από τα χαρακτηριστικά μάθησης (εδώ όσα είναι τα αντικείμενα που ανιχνεύονται από το YOLO, δηλαδή 80). Το επίπεδο εξόδου προσαρμόζεται ανάλογα με το πλήθος των κλάσεων, οπότε εδώ αποτελείται από 5 νευρώνες (1 για κάθε κλάση). Το πλήθος των κρυφών

επιπέδων, αλλά και το πλήθος των νευρώνων του καθενός, καθορίζονται έπειτα από εκτενή πειραματισμό.

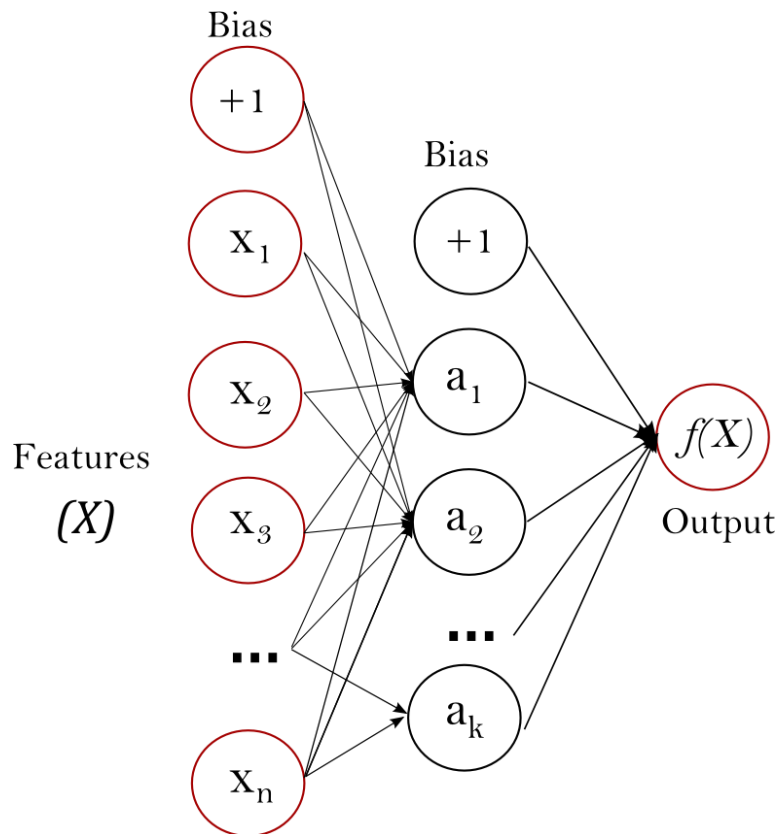
Σχετικά με την εκπαίδευση ενός νευρωνικού δικτύου χρησιμοποιείται συνήθως κάποια παραλλαγή του αλγορίθμου Οπισθοδιάδοσης (Backpropagation). Επίσης, χρησιμοποιούνται αλγόριθμοι γνωστοί ως optimizers ή solvers για τη μεταβολή των παραμέτρων του νευρωνικού δικτύου κατά την εκπαίδευση. Οι αλγόριθμοι αυτοί ελέγχουν τις τιμές μίας πληθώρας άλλων παραμέτρων κατά την εκπαίδευση και μπορεί να επηρεάσουν σημαντικά την απόδοση του νευρωνικού δικτύου. Ακόμη, ως loss function χρησιμοποιείται η Binary Cross Entropy, αφού μόνο αυτή παρέχεται από το Scikit-learn¹⁰.

Βάσει των παραπάνω, οι παράμετροι με τις οποίους γίνεται πειραματισμός με τα νευρωνικά δίκτυα MLP είναι οι εξής:

- **hidden_layer_sizes**, το πλήθος και μέγεθος (σε νευρώνες) των κρυφών επιπέδων. Δέχεται μία ακολουθία από τιμές με τη μορφή (a, b, c, ...), όπου το a καθορίζει το πλήθος των νευρώνων του πρώτου κρυφού επιπέδου, το b του δεύτερου, κ.ο.κ.
- **solver**, ο αλγόριθμος optimizer που χρησιμοποιείται κατά την εκπαίδευση. Οι διαθέσιμες επιλογές συμπεριλαμβάνουν τους αλγορίθμους Adam (μέσω της τιμής “adam”) και Stochastic Gradient Descent (μέσω της τιμής “sgd”).
- **alpha**, η παράμετρος του L2 regularization που χρησιμοποιείται κατά την εκπαίδευση.
- **max_iter**, ο μέγιστος αριθμός επαναλήψεων κατά την εκπαίδευση. Αν η απόδοση του αλγορίθμου συγκλίνει στη βέλτιστη προτού ξεπεραστεί ο μέγιστος αριθμός επαναλήψεων, η εκπαίδευση τερματίζεται. Εδώ τίθεται σε αρκετά υψηλή τιμή, ώστε να συγκλίνει πάντα ο αλγόριθμος προτού τερματιστεί η εκπαίδευση.
- **tol**, η ανεκτικότητα του αλγορίθμου στη μεταβολή της απόδοσης από επανάληψη σε επανάληψη, προτού θεωρηθεί ότι συγκλίνει.
- **learning_rate_init**, η αρχική τιμή του learning rate. Παίρνει σχετικά μικρές πραγματικές τιμές, με προεπιλεγμένη τιμή το 0.001.

Οι αλγόριθμοι ταξινόμησης που χρησιμοποιούνται κατά την αξιολόγηση της μεθοδολογίας, μαζί με τις σχετικές παραμέτρους, συνοψίζονται για λόγους συνάφειας στον Πίνακα 3.4.

¹⁰ https://scikit-learn.org/stable/modules/neural_networks_supervised.html



Σχήμα 3.8: Μία ενδεικτική αρχιτεκτονική ενός νευρωνικού δικτύου¹¹.

Πίνακας 3.4: Οι αλγόριθμοι ταξινόμησης και οι αντίστοιχες παράμετροι που χρησιμοποιούνται κατά τον πειραματισμό με το σύνολο δεδομένων.

| Αλγόριθμος Ταξινόμησης | Παράμετροι |
|------------------------|---|
| Dummy Classifier | strategy: {"stratified", "most_frequent", "uniform"} |
| Naive Bayes | - |
| Decision Tree | criterion: {"gini", "entropy"}, max_depth: {2, 3, ...}, class_weight: {None, "balanced"} |
| Random Forest | criterion: {"gini", "entropy"}, max_depth: {2, 3, ...}, class_weight: {None, "balanced", "balanced_subsample"}, n_estimators: {100, 1000, ...} |
| SVM | kernel: {"rbf"}, C: {1, 10, 1000, ...}, class_weight: {None, "balanced"} |
| MLP | solver: {"adam", "sgd"}, alpha: { 10^{-4} , 10^{-5} }, hidden_layer_sizes: {(10, 10), (30, 20, 15), (100, 50, 25), (200, 100), (200, 100, 50, 25)}, max_iter: {2500}, tol={ 10^{-6} , 10^{-8} }, learning_rate_init={0.01, 0.1} |

¹¹ https://scikit-learn.org/stable/modules/neural_networks_supervised.html#multi-layer-perceptron

4. Αποτελέσματα Αξιολόγησης Αλγορίθμων

Στο κεφάλαιο αυτό, αναλύεται εκτενώς ο πειραματισμός που λαμβάνει χώρα με τους αλγορίθμους ταξινόμησης, ώστε να αξιολογηθεί η αποτελεσματικότητα της μεθοδολογίας. Για την αξιολόγηση χρησιμοποιούνται διάφορες μετρικές οι οποίες αφορούν είτε στη binary ταξινόμηση είτε στη multi-label ταξινόμηση, για τις οποίες παρέχεται ο ορισμός. Στη συνέχεια, και σχετικά με τα αποτελέσματα, γίνεται αναλυτικός σχολιασμός και σύγκριση μεταξύ των διαφόρων αλγορίθμων και παραμέτρων.

4.1 Μετρικές Αξιολόγησης

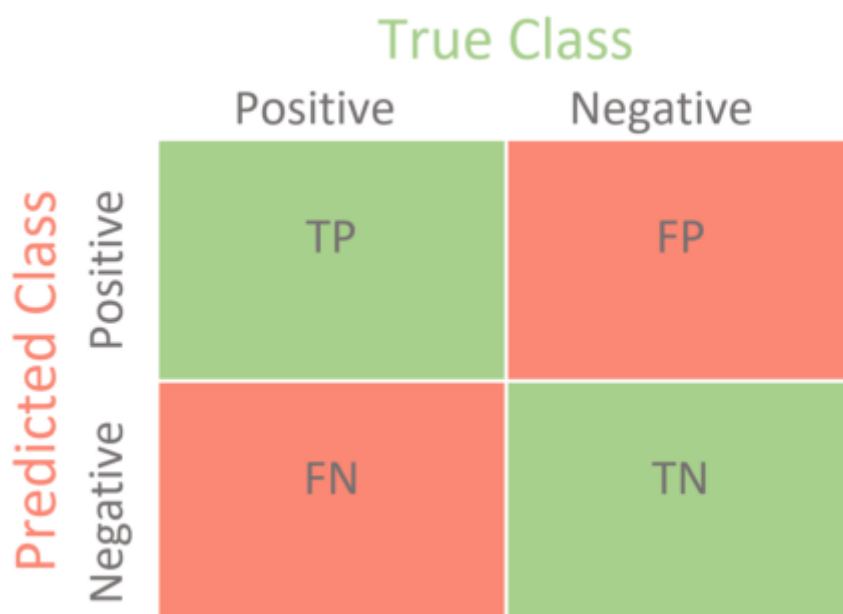
Οι μετρικές αξιολόγησης, ιδιαίτερα όσες έχουν να κάνουν με τη multi-label ταξινόμηση, είναι ένα σύνθετο ζήτημα και σίγουρα ιδιαίτερα σημαντικό, αφού σε αυτές βασίζεται η αξιολόγηση όλης της μεθοδολογίας. Στη συνέχεια, παρουσιάζεται το θεωρητικό υπόβαθρο από μερικές από τις πιο διαδεδομένες μετρικές, όπως το accuracy, το precision, το recall, κ.α., που έχουν να κάνουν με binary ταξινόμηση, ενώ στη συνέχεια παρουσιάζονται πιο σύνθετες που σχετίζονται με τη multi-label ταξινόμηση.

Αρχικά, έστω X ο διδιάστατος πίνακας με τα χαρακτηριστικά μάθησης του συνόλου δεδομένων, με το πλήθος των γραμμών να είναι $n_{samples}$, ενώ των στηλών $n_{features}$. Αντίστοιχα, έστω $Y = \{0, 1\}^{n_{samples} \times n_{classes}}$ ο διδιάστατος πίνακας που αναπαριστά τις τιμές των κλάσεων του συνόλου δεδομένων, με το πλήθος των γραμμών να είναι και πάλι $n_{samples}$, ενώ το πλήθος των στηλών $n_{classes} = |\{Drama, Comedy, Romance, Action, Thriller\}| = 5$.

4.1.1 Binary Ταξινόμηση

Στην περίπτωση της binary ταξινόμησης, οι μετρικές που χρησιμοποιούνται ευρέως και αναλύονται παρακάτω είναι οι εξής: accuracy, precision, recall και F-Score (Hossin & Sulaiman, 2015). Μιας κι εδώ οι μετρικές αφορούν μόνο μία κλάση, ο συμβολισμός Y αναπαριστά μία μόνο στήλη του αρχικού πίνακα, όπως ορίστηκε παραπάνω.

Για όλες τις μετρικές, ιδιαίτερα σημαντικές είναι οι έννοιες των True Positives (TP), True Negatives (TN), False Positives (FP) και False Negatives (FN). Συγκεκριμένα, οι έννοιες αυτές αποτελούνται από δύο μέρη: True/False για το πρώτο και Positive/Negative για το δεύτερο. Το Positive/Negative υποδηλώνει την πρόβλεψη ενός αλγορίθμου ταξινόμησης για αυτό το δείγμα, ενώ το True/False την ορθότητα της πρόβλεψης.



Σχήμα 4.1: Η δομή ενός confusion matrix, που χρησιμοποιείται για τον χαρακτηρισμό των προβλέψεων ενός αλγορίθμου ταξινόμησης¹².

Έτσι, έστω πως μελετά κανείς την κατηγορία Thriller (οι υπόλοιπες κλάσεις αγνοούνται, μιας κι εδώ μελετάται μόνο η binary ταξινόμηση) και ισχύει πως θετική τιμή (δηλ. 1) της κλάσης δείχνει ότι μία ταινία ανήκει σε αυτό το είδος. Τότε, για μία ταινία που όντως ανήκει στην κλάση Thriller, μία θετική πρόβλεψη θεωρείται TP, αφού η πρόβλεψη είναι θετική (άρα Positive) και είναι και σωστή (άρα True). Βάσει αυτών των εννοιών, το συνολικό πλήθος της ορθότητας των προβλέψεων συγκεντρώνεται συνήθως σε έναν πίνακα σύγχυσης (confusion matrix), όπως φαίνεται στο Σχήμα 4.1.

Το accuracy είναι μάλλον η απλούστερη μετρική, αφού αναπαριστά το ποσοστό των σωστών προβλέψεων. Έτσι, βάσει του confusion matrix, ορίζεται ως $accuracy = \frac{TP + TN}{TP + TN + FP + FN}$, ενώ, αν \hat{y}_i η πρόβλεψη ενός αλγορίθμου για το δείγμα i , ο πιο μαθηματικός ορισμός του accuracy είναι $accuracy = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i)$, όπου $1(x)$ ισούται με 1 αν το x είναι λογικά true, αλλιώς ισούται με 0. Ένα μειονέκτημα του accuracy είναι πως, ως μετρική, δεν είναι ιδιαίτερα χρήσιμη για περιπτώσεις όπου τα δείγματα είναι ανομοιόμορφα κατανομημένα στις τιμές μίας κλάσης. Για παράδειγμα, αν το 70% των ταινιών ανήκουν στο είδος Thriller και ο αλγόριθμος ταξινόμησης προβλέπει πάντα πως μία ταινία είναι Thriller, το accuracy θα είναι 70% (αρκετά υψηλό), όμως για τη μία κλάση δεν θα είναι ποτέ σωστή η πρόβλεψη.

Μία άλλη μετρική είναι το precision, το οποίο δείχνει πόσες από τις θετικές προβλέψεις είναι σωστές. Δηλαδή, είναι η αναλογία των σωστών θετικών προβλέψεων προς τις συνολικές

¹² <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>

θετικές προβλέψεις ή $precision = \frac{TP}{TP + FP}$. Μία αντίστοιχη μετρική είναι και το recall, το οποίο δείχνει το πόσα θετικά δείγματα προέβλεψε σωστά ο αλγόριθμος ταξινόμησης. Ορίζεται ως η αναλογία μεταξύ των σωστών θετικών προβλέψεων προς το πλήθος των συνολικών θετικών δειγμάτων ή $recall = \frac{TP}{TP + FN}$.

Μιας και δεν είναι σπάνιο οι μετρικές των precision και recall να απέχουν αρκετά μεταξύ τους, μία μετρική που χρησιμοποιείται συχνά, η οποία συνδυάζει τις δύο προηγούμενες, είναι το F-Score. Το F-Score ορίζεται ως ο αρμονικός μέσος μεταξύ του precision και του recall ή $F-Score = 2 \frac{precision \cdot recall}{precision + recall}$ και συχνά είναι πιο αντιπροσωπευτική μετρική από το accuracy, ειδικά για imbalanced δεδομένα. Για τον λόγο αυτόν, εδώ χρησιμοποιείται ως μία από τις κυριότερες μετρικές για την αξιολόγηση των αλγορίθμων.

Ενδιαφέρον παρουσιάζει το γεγονός ότι οι τιμές των TP, TN, FP και FN και κατ' επέκταση το precision, recall και F-Score βασίζονται στον ορισμό της θετικής κλάσης. Συνήθως, η θετική κλάση υποδηλώνει το λογικό true, οπότε, στην περίπτωση των ταινιών, ότι μία ταινία ανήκει όντως σε ένα συγκεκριμένο είδος. Όμως, έτσι οι μετρικές δεν υποδηλώνουν κάτι για την αρνητική κλάση, δηλαδή το πόσο αποτελεσματικά ένας αλγόριθμος προβλέπει πως μία ταινία δεν ανήκει σε ένα είδος. Για τον λόγο αυτόν, συχνά υπολογίζονται οι μετρικές διπλά: μία φορά θεωρώντας ως θετική τιμή το λογικό true και μία δεύτερη φορά θεωρώντας ως θετική τιμή το λογικό false. Έτσι, οι μετρικές είναι αντιπροσωπευτικές και για τις δύο τιμές κλάσεων και συχνά χρησιμοποιούνται οι μέσοι όροι των δύο τιμών της εκάστοτε μετρικής για τη συνολική αξιολόγηση. Έτσι κι εδώ, υπολογίζονται τα precision, recall και F-Score για κάθε τιμή της κλάσης και παρατίθεται κι ο μέσος όρος τους.

4.1.2 Multi-label Ταξινόμηση

Μία βασική διαφορά μεταξύ της binary ταξινόμησης και της multi-label ταξινόμησης είναι πως, στην πρώτη περίπτωση, οι προβλέψεις ενός αλγορίθμου θα είναι είτε απόλυτα σωστές, είτε απόλυτα λάθος, ενώ στη δεύτερη περίπτωση, οι προβλέψεις μπορεί να είναι μερικώς σωστές (π.χ., να είναι σωστές οι προβλέψεις στα 3 από τα 5 είδη ταινιών). Για τον λόγο αυτόν, έχουν προταθεί διάφορες μετρικές, πιο σύνθετες (Sorower, 2010) για την αξιολόγηση αλγορίθμων ταξινόμησης σε multi-label προβλήματα, οι οποίες περιγράφονται στη συνέχεια.

Ένας σχετικά απλός τρόπος να παρακάμψει κανείς το γεγονός ότι κάποιες προβλέψεις είναι μερικώς σωστές είναι να τις αγνοήσει. Για την ακρίβεια, το Exact Match Ratio είναι μία μετρική που υπολογίζει το ποσοστό των απόλυτα σωστών προβλέψεων προς όλες τις προβλέψεις, θεωρώντας τις μερικώς σωστές ως λανθασμένες. Έτσι, ορίζεται ως $Exact Match Ratio = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} 1(Y_i = \hat{Y}_i)$, το οποίο ως εξίσωση μοιάζει με αυτήν του accuracy, όμως διαφέρει η συνθήκη του $Y_i = \hat{Y}_i$, αφού εδώ τα Y, \hat{Y} είναι διανύσματα κι όχι απλές τιμές.

Μία ακόμη μετρική που έχει προταθεί για την αξιολόγηση multi-label προβλημάτων είναι το Hamming Loss. Το Hamming Loss λαμβάνει υπόψη τόσο τις λανθασμένες προβλέψεις (FP) όσο και τις ελλιπείς προβλέψεις (FN), κανονικοποιημένες ως προς το συνολικό πλήθος των κλάσεων και των δειγμάτων. Έτσι, το Hamming Loss ορίζεται ως:

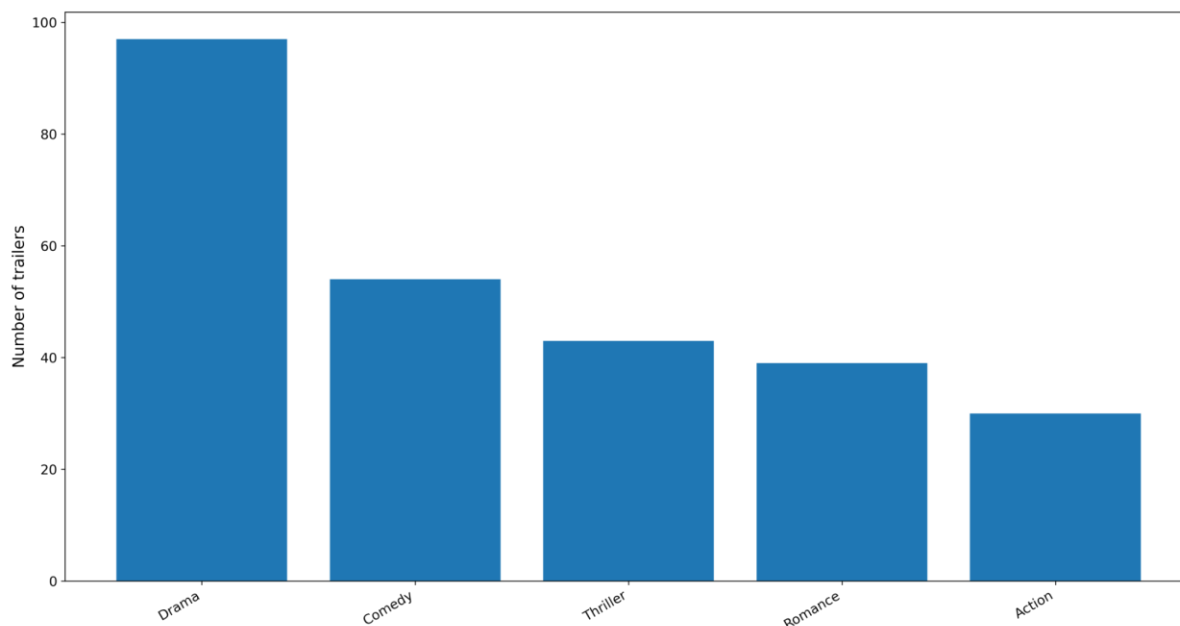
$$\text{Hamming Loss} = \frac{1}{n_{\text{samples}} n_{\text{classes}}} \sum_{i=1}^{n_{\text{samples}}} \sum_{l=1}^{n_{\text{classes}}} [1((\hat{Y}_i = 1) \wedge (Y_i == 0)) + 1((\hat{Y}_i == 0) \wedge (Y_i = 1))].$$

Αξίζει να σημειωθεί πως όσο χαμηλότερη είναι η τιμή του Hamming Loss, τόσο καλύτερο για την απόδοση του αλγορίθμου.

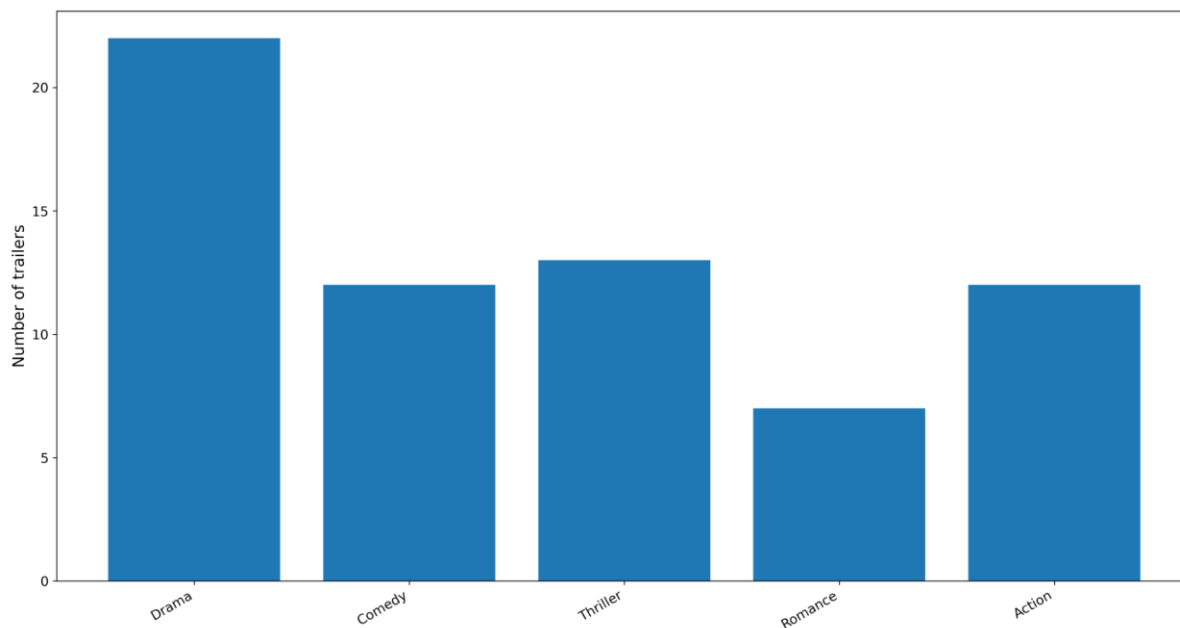
Ακόμη, χρησιμοποιούνται και πάλι οι μετρικές Precision, Recall και F-Score, όμως εδώ υπολογίζονται με διαφορετικό τρόπο. Για την ακρίβεια, για κάθε δείγμα κατασκευάζεται ένα confusion matrix για τις 5 προβλέψεις των ειδών στα οποία πιθανώς αυτό να ανήκει. Με βάση αυτόν, υπολογίζονται οι μετρικές Precision, Recall και F-Score. Η διαδικασία αυτή επαναλαμβάνεται για κάθε δείγμα και υπολογίζεται ο μέσος όρος κάθε μετρικής. Ο τύπος υπολογισμού αυτών των μετρικών συχνά αναφέρεται ως sample-wise average και χρησιμοποιείται κι εδώ για τη συνολική αξιολόγηση του συστήματος των ταξινομητών.

4.2 Αποτελέσματα

Για την αξιολόγηση των αλγορίθμων, το σύνολο δεδομένων ταξινόμησης συνήθως διαχωρίζεται σε δύο (τουλάχιστον) τμήματα: το training set και το test set. Το πρώτο χρησιμοποιείται για την εκπαίδευση του αλγορίθμου, ενώ το δεύτερο για την αξιολόγηση. Το βήμα αυτό είναι σημαντικό καθώς ο στόχος της αξιολόγησης είναι να εξεταστεί αν το μοντέλο που παράγεται από έναν αλγόριθμο είναι ικανό να παράγει σωστές προβλέψεις για δείγματα που δεν έχει ξαναδεί. Εδώ, το training set αποτελείται από το 80% (167) των δειγμάτων του αρχικού συνόλου, ενώ το test set από το υπόλοιπο 20% (42). Είναι συνετό η κατανομή των δειγμάτων και στο training set και στο test set να είναι όσο γίνεται όμοια, αν και αυτό δεν είναι εύκολο, δεδομένου ότι ένα δείγμα ανήκει σε πολλαπλές κλάσεις και ο διαχωρισμός πραγματοποιείται τυχαία. Με τον διαχωρισμό που πραγματοποιείται εδώ, οι κατανομές που προκύπτουν για τα δύο υποσύνολα παρουσιάζονται στα Σχήματα 4.2 και 4.3.



Σχήμα 4.2: Η κατανομή των δειγμάτων στις 5 διαφορετικές κλάσεις στο training set.



Σχήμα 4.3: Η κατανομή των δειγμάτων στις 5 διαφορετικές κλάσεις στο test set.

4.2.1 Αξιολόγηση στο Multi-Label Πρόβλημα Ταξινόμησης

Έπειτα, ακολουθεί ο πειραματισμός με τους διάφορους αλγόριθμους ταξινόμησης και τις αντίστοιχες παραμέτρους τους. Για την ακρίβεια, για κάθε αλγόριθμο πραγματοποιούνται πολλαπλές δοκιμές, με την εκπαίδευση να λαμβάνει χώρα στο training set και την αξιολόγηση στο test set, ώστε να εντοπιστεί ο συνδυασμός των παραμέτρων που επιφέρει την υψηλότερη απόδοση. Μιας και χρησιμοποιούνται πολλές μετρικές, ο αλγόριθμος με την υψηλότερη απόδοση επιλέγεται με κύριο κριτήριο το F-Score που προκύπτει ως το sample average. Αξίζει να σημειωθεί πως επιλέγεται αυτό το κριτήριο ως μετρική για τη συνολική απόδοση της multi-label ταξινόμησης κι όχι για κάποια επιμέρους κλάση, ενώ το F-Score, όπως έχει ήδη

αναφερθεί, συνδυάζει τις μετρικές του Recall και του Precision, οπότε είναι πιο αντιπροσωπευτική.

Αρχικά, παρέχεται ένας πίνακας με τη συνολική απόδοση κάθε αλγορίθμου ανεξάρτητα του είδους της ταινίας για τη σύγκρισή τους στην αντιμετώπιση του multi-label προβλήματος ταξινόμησης. Επιπλέον, ανεξάρτητα του είδους μιας ταινίας και του αλγορίθμου ταξινόμησης, έπονται κάποιοι πίνακες με κοινή δομή, οι οποίοι περιέχουν την απόδοση κάθε αλγορίθμου για την ταξινόμηση του test set σε κάθε είδος ξεχωριστά, παρουσιάζοντας τις τιμές των μετρικών Precision, Recall και F-Score, οι οποίες υπολογίζονται με το sample-wise average, όπως έχει προαναφερθεί. Σε όλους τους πίνακες εμφανίζονται με έντονη γραμματοσειρά τα υψηλότερα των αποτελεσμάτων για κάθε μετρική αξιολόγησης.

Πίνακας 4.1: Η συνολική απόδοση των αλγορίθμων.

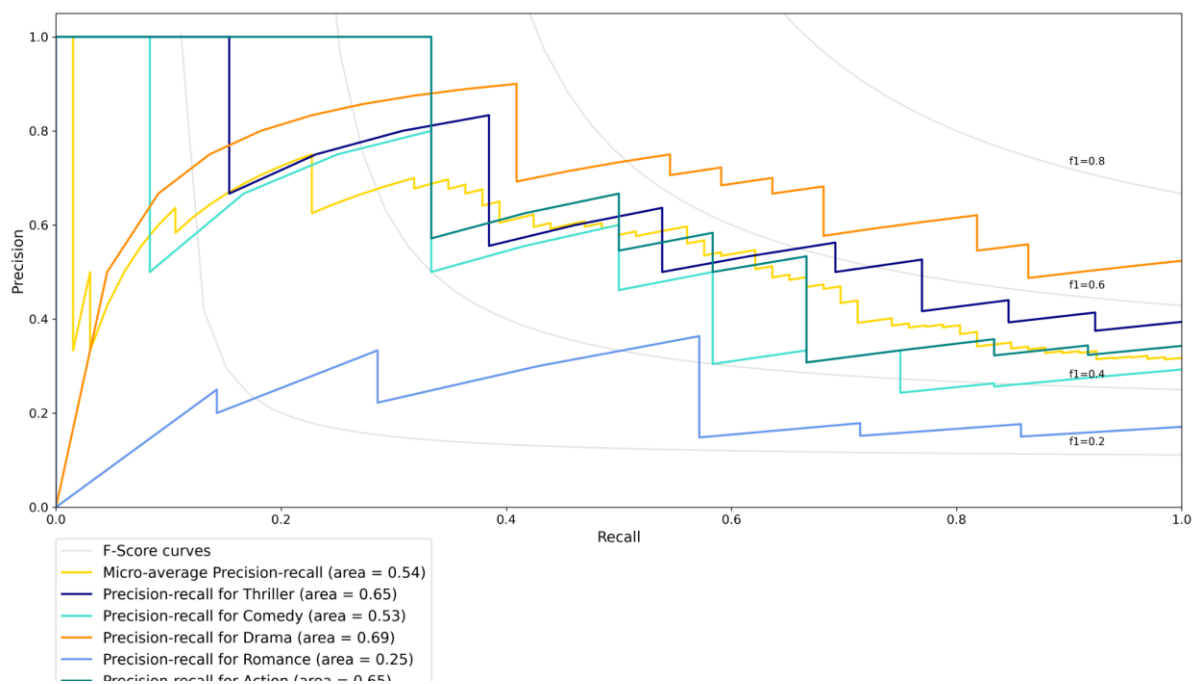
| Αλγόριθμος | Precision | Recall | F-Score | Exact Match Ratio | Hamming Loss |
|------------------|------------|------------|------------|-------------------|---------------|
| Dummy Classifier | 43% | 40% | 37% | 0.1667 | 0.4142 |
| Naive Bayes | 43% | 68% | 47% | 0.0714 | 0.4143 |
| Decision Tree | 54% | 51% | 49% | 0.2143 | 0.2571 |
| Random Forest | 62% | 56% | 56% | 0.3333 | 0.2476 |
| SVM | 59% | 50% | 51% | 0.2381 | 0.2667 |
| MLP | 50% | 60% | 51% | 0.1904 | 0.3285 |

Με στόχο τη γενικότερη σύγκριση της απόδοσης των αλγορίθμων ανεξάρτητα από το είδος μιας ταινίας, παρουσιάζεται ο Πίνακας 4.1, ο οποίος περιλαμβάνει τις μετρικές αξιολόγησης για multi-label προβλήματα που έχουν ήδη παρουσιαστεί, για τη δημιουργία μιας ευρύτερης εικόνας για κάθε αλγόριθμο ξεχωριστά. Από αυτόν φαίνεται πως ο αλγόριθμος Random Forest υπερτερεί στο σύνολο των μετρικών, με εξαίρεση το Recall, στο οποίο την υψηλότερη απόδοση παρουσιάζει ο Naive Bayes. Ωστόσο, αξίζει να σημειωθεί ότι, σε σχέση με τον αλγόριθμο Dummy Classifier, ο οποίος χρησιμοποιείται ως μέτρο σύγκρισης, υπάρχει μία σημαντική βελτίωση. Το γεγονός αυτό είναι ενθαρρυντικό για τη μεθοδολογία αφού, για παράδειγμα, το F-Score διαφέρει κατά 19%.

Έχοντας παρουσιάσει τα αποτελέσματα των αλγορίθμων, είναι απαραίτητο να αναφερθούν και οι τελικές παράμετροι που επιλέχθηκαν και οι οποίες συνοψίζονται στον Πίνακα 4.2 για κάθε αλγόριθμο.

Πίνακας 4.2: Οι παράμετροι των αλγορίθμων ταξινόμησης που επιτυγχάνουν τα καλύτερα αποτελέσματα.

| Αλγόριθμος Ταξινόμησης | Παράμετροι |
|------------------------|---|
| Dummy Classifier | strategy: "stratified" |
| Naive Bayes | - |
| Decision Tree | criterion: "gini", max_depth: 11, class_weight: None |
| Random Forest | criterion: "entropy", max_depth: 8, class_weight: "balanced_subsample", n_estimators: 1000 |
| SVM | kernel: "rbf", C: 1000, class_weight: None |
| MLP | layer_size: (30, 20, 15), solver: "adam", alpha: 10^{-5} , hidden_layer_sizes: (30, 20, 15), max_iter: 2500, tol= 10^{-8} , learning_rate_init=0.01 |



Σχήμα 4.4: Η καμπύλη Precision-Recall που προκύπτει από τον αλγόριθμο SVM για κάθε κλάση.

Ενδιαφέρον παρουσιάζει το γεγονός πως, αν και στον Πίνακα 4.1 δίνεται έμφαση σε έναν αλγόριθμο που πετυχαίνει κατά μέσο όρο τις καλύτερες τιμές στις σχετικές μετρικές, θα μπορούσε να γίνει η σύνθεση ενός συστήματος με διαφορετικούς αλγορίθμους ανά κλάση. Για την ακρίβεια, δεδομένου ότι το multi-label πρόβλημα ταξινόμησης μετατρέπεται σε μία σειρά από ανεξάρτητα προβλήματα binary ταξινόμησης, θα μπορούσε κανείς να επιλέξει τον αλγόριθμο με τις καλύτερες μετρικές ανά binary πρόβλημα ταξινόμησης, δηλαδή ανά είδος ταινίας. Με τον τρόπο αυτόν, μπορούν να επιτευχθούν ακόμη καλύτερα αποτελέσματα από

αυτά που επιτυγχάνει ο Random Forest. Αυτό, μάλιστα, επιβεβαιώνεται και στη συνέχεια, αφού ο Random Forest δεν φαίνεται να επιτυγχάνει τις καλύτερες αποδόσεις σε κανένα επιμέρους είδος ταινίας, αλλά συνολικά έχει σταθερά υψηλές αποδόσεις, οπότε και καταλήγει να είναι ο πιο αποδοτικός αλγόριθμος κατά μέσο όρο.

Ένας διαφορετικός τρόπος αξιολόγησης της απόδοσης ενός αλγορίθμου ταξινόμησης, ειδικά σε προβλήματα που παρουσιάζεται imbalance, είναι η καμπύλη Precision-Recall, η οποία παρουσιάζεται στο Σχήμα 4.4 για τον αλγόριθμο SVM για κάθε κλάση. Τα διαγράμματα αυτού του τύπου απεικονίζουν τον τρόπο που μεταβάλλεται το Precision και το Recall καθώς μεταβάλλεται η ευαισθησία του αλγορίθμου στο να παράγει μία θετική πρόβλεψη. Όπως φαίνεται κι από το συγκεκριμένο σχήμα, αν για το σύστημα ταξινόμησης θεωρηθεί σημαντικό να διαθέτει μέσο Recall κοντά στο 100%, τότε ο αλγόριθμος θα παράγει αρκετές λάθος προβλέψεις, αφού το Precision θα είναι μικρότερο του 40%.

4.2.2 Αξιολόγηση στα Επιμέρους Binary Προβλήματα Ταξινόμησης

Τα αποτελέσματα που αφορούν τις ταινίες που χαρακτηρίζονται ως Drama και η απόδοση των αλγορίθμων ταξινόμησης για αυτές απεικονίζονται στον Πίνακα 4.3. Πιο συγκεκριμένα, φαίνεται πως ο αλγόριθμος με την καλύτερη απόδοση είναι ο αλγόριθμος Decision Tree με 69% σε όλες τις μετρικές αξιολόγησης ενώ, αντίθετα, τη χαμηλότερη απόδοση παρουσιάζει ο αλγόριθμος Naive Bayes. Ωστόσο, η απόδοση όλων των αλγορίθμων δεν φαίνεται να έχει ιδιαίτερες αποκλίσεις, ενώ η απόδοση του Recall σε όλους τους αλγορίθμους, πλην του Decision Tree, για την κλάση True είναι πιο υψηλή. Αυτό το γεγονός υποδηλώνει ότι υπάρχει μια προτίμηση στην κλάση True.

Πίνακας 4.3: Η απόδοση των αλγορίθμων για τις ταινίες Drama.

| Αλγόριθμος | Κλάση | Precision | Recall | F-Score |
|------------------|---------|-----------|--------|---------|
| Dummy Classifier | False | 59% | 50% | 54% |
| | True | 60% | 68% | 64% |
| | Average | 59% | 59% | 59% |
| Naive Bayes | False | 58% | 35% | 44% |
| | True | 57% | 77% | 65% |
| | Average | 57% | 56% | 55% |
| Decision Tree | False | 67% | 70% | 68% |
| | True | 71% | 68% | 70% |

| | | | | |
|---------------|---------|------------|------------|------------|
| | Average | 69% | 69% | 69% |
| Random Forest | False | 73% | 40% | 52% |
| | True | 61% | 86% | 72% |
| | Average | 67% | 63% | 62% |
| SVM | False | 59% | 50% | 54% |
| | True | 60% | 68% | 64% |
| | Average | 59% | 59% | 59% |
| MLP | False | 67% | 30% | 41% |
| | True | 58% | 86% | 69% |
| | Average | 62% | 58% | 55% |

Αναφορικά με την απόδοση των χρησιμοποιηθέντων αλγορίθμων για το είδος Comedy, αυτή απεικονίζεται στον Πίνακα 4.4. Γίνεται φανερό πως ο αλγόριθμος με τα υψηλότερα αποτελέσματα είναι ο Naive Bayes, ενώ εν συγκρίσει με τον αλγόριθμο Dummy Classifier εντοπίζεται μια ποσοστιαία διαφορά της τάξης του 37%, κάτι που επιβεβαιώνει την αποτελεσματικότητα της μεθοδολογίας. Επιπλέον, σε αντίθεση με το είδος Drama, τόσο το Precision, όσο και το Recall παρουσιάζουν υψηλότερες τιμές στην κλάση False παρά στην True.

Πίνακας 4.4: Η απόδοση των αλγορίθμων για τις ταινίες Comedy.

| Αλγόριθμος | Κλάση | Precision | Recall | F-Score |
|------------------|---------|-----------|--------|---------|
| Dummy Classifier | False | 62% | 53% | 57% |
| | True | 12% | 17% | 14% |
| | Average | 37% | 35% | 36% |
| Naive Bayes | False | 84% | 87% | 85% |
| | True | 64% | 58% | 61% |

| | | | | |
|---------------|---------|------------|------------|------------|
| | Average | 74% | 73% | 73% |
| Decision Tree | False | 80% | 80% | 80% |
| | True | 50% | 50% | 50% |
| | Average | 65% | 65% | 65% |
| Random Forest | False | 70% | 90% | 79% |
| | True | 40% | 15% | 22% |
| | Average | 55% | 53% | 51% |
| SVM | False | 79% | 87% | 83% |
| | True | 56% | 42% | 48% |
| | Average | 67% | 64% | 65% |
| MLP | False | 76% | 87% | 81% |
| | True | 50% | 33% | 40% |
| | Average | 63% | 60% | 61% |

Τα αποτελέσματα για τις ταινίες που χαρακτηρίζονται Romance απεικονίζονται στον Πίνακα 4.5. Ξεκινώντας από τον αλγόριθμο Dummy Classifier, φαίνεται πως η απόδοσή του για την κλάση True, ανεξαρτήτως μετρικής αξιολόγησης είναι η χειρότερη δυνατή, καθώς δε ξεπερνά το 0%. Αυτό μπορεί να οφείλεται στο γεγονός ότι μόλις 7 από τα δείγματα ανήκουν στην πραγματική κλάση True.

Παράλληλα και εν αντιθέσει με τα αποτελέσματα των προηγούμενων ειδών, την υψηλότερη απόδοση σε κάθε μετρική αξιολόγησης την παρουσιάζει διαφορετικός αλγόριθμος. Για την ακρίβεια, ο αλγόριθμος Random Forest παρουσιάζει το υψηλότερο Precision με 93%, ενώ ο αλγόριθμος Decision Tree έχει τα πιο ικανοποιητικά αποτελέσματα στις υπόλοιπες μετρικές. Ακόμη, αξίζει να αναφερθεί πως το Precision έχει υψηλότερες τιμές για την κλάση True. Επιπλέον, αξιοσημείωτο είναι το γεγονός πως, παρ' ότι χρησιμοποιείται η τιμή "balanced" για την παράμετρο class_weights του Random Forest, υπάρχει έντονη διαφορά μεταξύ της τιμής του Recall για την κλάση False (100%) σε σχέση με την κλάση True (14%), οπότε η παράμετρος δε φαίνεται να είναι ιδιαίτερα αποτελεσματική σε όλες τις περιπτώσεις.

Πίνακας 4.5: Η απόδοση των αλγορίθμων για τις ταινίες Romance.

| Αλγόριθμος | Κλάση | Precision | Recall | F-Score |
|------------------|---------|------------|------------|------------|
| Dummy Classifier | False | 79% | 74% | 76% |
| | True | 0% | 0% | 0% |
| | Average | 39% | 37% | 38% |
| Naive Bayes | False | 85% | 49% | 62% |
| | True | 18% | 57% | 28% |
| | Average | 52% | 53% | 45% |
| Decision Tree | False | 88% | 86% | 87% |
| | True | 38% | 43% | 40% |
| | Average | 63% | 64% | 63% |
| Random Forest | False | 85% | 100% | 92% |
| | True | 100% | 14% | 25% |
| | Average | 93% | 57% | 59% |
| SVM | False | 84% | 89% | 86% |
| | True | 20% | 14% | 17% |
| | Average | 52% | 51% | 51% |
| MLP | False | 81% | 71% | 76% |
| | True | 9% | 14% | 11% |
| | Average | 45% | 43% | 43% |

Στον Πίνακα 4.6. παρουσιάζονται τα αποτελέσματα των αλγορίθμων για τις ταινίες Action. Ένα ενδιαφέρον στοιχείο είναι πως ο αλγόριθμος Decision Tree και ο αλγόριθμος Random Forest παρουσιάζουν ακριβώς την ίδια απόδοση, ενώ ο αλγόριθμος SVM έχει την υψηλότερη

απόδοση για τις μετρικές Recall και F-score. Επιπρόσθετα, τόσο το Precision όσο και το Recall έχουν υψηλότερες τιμές στην κλάση True. Αν και ο αλγόριθμος Dummy Classifier έχει την υψηλότερη απόδοση στις μετρικές για αυτό το είδος ταινιών και πάλι η χρησιμοποιούμενη μεθοδολογία φαίνεται να υπερτερεί, παρουσιάζοντας υψηλότερα ποσοστά.

Πίνακας 4.6: Η απόδοση των αλγορίθμων για τις ταινίες Action.

| Αλγόριθμος | Κλάση | Precision | Recall | F-Score |
|------------------|---------|------------|------------|------------|
| Dummy Classifier | False | 74% | 83% | 78% |
| | True | 38% | 25% | 30% |
| | Average | 56% | 54% | 54% |
| Naive Bayes | False | 72% | 60% | 65% |
| | True | 29% | 42% | 34% |
| | Average | 51% | 51% | 50% |
| Decision Tree | False | 80% | 93% | 86% |
| | True | 71% | 42% | 53% |
| | Average | 76% | 68% | 69% |
| Random Forest | False | 80% | 93% | 86% |
| | True | 71% | 42% | 53% |
| | Average | 76% | 68% | 69% |
| SVM | False | 82% | 90% | 86% |
| | True | 67% | 50% | 57% |
| | Average | 74% | 70% | 71% |
| MLP | False | 82% | 77% | 79% |
| | True | 50% | 58% | 54% |

| | | | | |
|--|---------|-----|-----|-----|
| | Average | 66% | 68% | 67% |
|--|---------|-----|-----|-----|

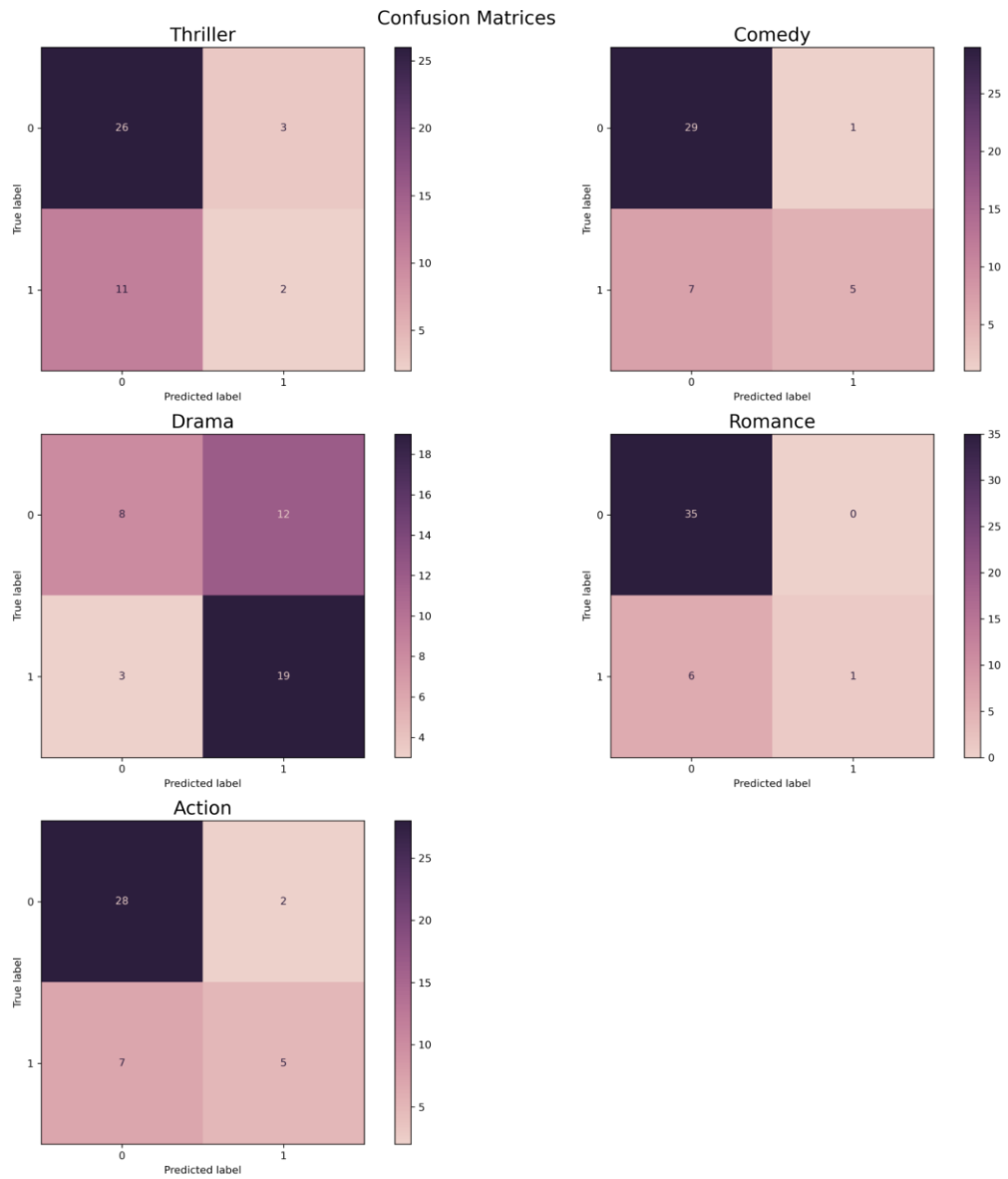
Τέλος, και αναφορικά με τις ταινίες Thriller, η απόδοση των αλγορίθμων περιέχεται στον Πίνακα 4.7. Όπως γίνεται φανερό, ο αλγόριθμος SVM έχει την υψηλότερη μέση απόδοση, πετυχαίνοντας 81% precision, 68% recall και 69% F-score. Επιπλέον, σε σύγκριση με τον αλγόριθμο Dummy Classifier, ο οποίος έχει μέτρια απόδοση, λίγο παραπάνω από 50%, οι υπόλοιποι αλγόριθμοι παρουσιάζουν σημαντική βελτίωση. Ακόμη, παρατηρείται και εδώ πως ο Random Forest έχει υψηλότερες τιμές στο Recall για την κλάση False (90%) απ' ότι για την True (15%), παρά τη χρήση της παραμέτρου class_weights.

Πίνακας 4.7: Η απόδοση των αλγορίθμων για τις ταινίες Thriller.

| Αλγόριθμος | Κλάση | Precision | Recall | F-Score |
|------------------|---------|-----------|--------|---------|
| Dummy Classifier | False | 71% | 76% | 73% |
| | True | 36% | 31% | 33% |
| | Average | 54% | 53% | 53% |
| Naive Bayes | False | 80% | 41% | 55% |
| | True | 37% | 77% | 50% |
| | Average | 59% | 59% | 52% |
| Decision Tree | False | 76% | 90% | 83% |
| | True | 62% | 38% | 48% |
| | Average | 69% | 64% | 65% |
| Random Forest | False | 70% | 90% | 79% |
| | True | 40% | 15% | 22% |
| | Average | 55% | 53% | 51% |
| SVM | False | 78% | 97% | 76% |
| | True | 83% | 38% | 53% |

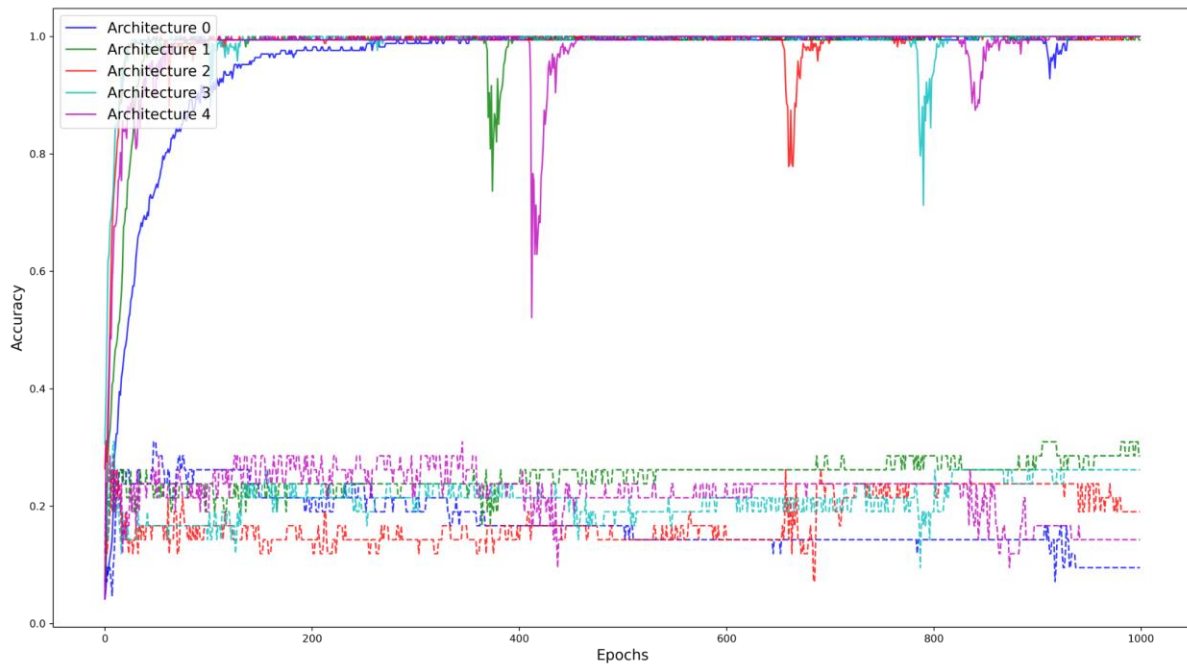
| | | | | |
|-----|---------|------------|------------|------------|
| | Average | 81% | 68% | 69% |
| MLP | False | 76% | 86% | 81% |
| | True | 56% | 38% | 45% |
| | Average | 66% | 62% | 63% |

Στο Σχήμα 4.5 παρουσιάζονται τα confusion matrices για κάθε είδος ταινίας που εξάγονται από τον αλγόριθμο Random Forest. Παρατηρείται ότι, στο σύνολο των ειδών, εκτός του Drama, υπάρχει ένα μεγάλο πλήθος από TNs. Αντίθετα, στο είδος Drama συναντώνται αρκετά TPs και λιγότερα TNs. Το γεγονός αυτό υποδηλώνει πως ο αλγόριθμος Random Forest τείνει να δείχνει προτίμηση σε μία κλάση, πιθανότατα λόγω imbalance. Αυτό, άλλωστε, επιβεβαιώνεται κι από το γεγονός ότι για κάθε είδος, εκτός του Drama, παρατηρείται πως γίνονται το πολύ 7 σωστές προβλέψεις για τη θετική κλάση και, μάλιστα, στο Romance μόνο μία πρόβλεψη για τη θετική κλάση είναι σωστή. Ακόμη, τα FNs παρουσιάζουν παρόμοιες τιμές στις κλάσεις Comedy, Romance και Action. Η συμπεριφορά αυτή είναι ενδιαφέρουσα, δεδομένου ότι στον συγκεκριμένο αλγόριθμο χρησιμοποιείται και η παράμετρος `class_weight` με τιμή `balanced_subsamples`, με στόχο την αντιμετώπιση του imbalance.



Σχήμα 4.5: Τα confusion matrices που προκύπτουν από την αξιολόγηση του Random Forest για κάθε είδος ταινίας.

4.2.3 Εκπαίδευση και Παραμετροποίηση του Νευρωνικού Δικτύου



Σχήμα 4.6: Η απόδοση των αλγορίθμων για τις διαφορετικές αρχιτεκτονικές MLP. Με τις συμπαγείς γραμμές εμφανίζεται το accuracy κατά την εκπαίδευση, ενώ με τις διακεκομμένες το accuracy κατά την αξιολόγηση.

Σε αυτό το σημείο παρουσιάζεται η διαδικασία εκπαίδευσης και παραμετροποίησης του MLP, ώστε να επιλεγεί η τελική αρχιτεκτονική και οι σχετικές παράμετροι. Για την αξιολόγηση κάθε αρχιτεκτονικής, χρησιμοποιούνται οι καμπύλες του Accuracy που υπολογίζονται σε κάθε επανάληψη κατά την εκπαίδευση για το training set και για το test set. Έτσι, μπορεί κανείς να παρατηρήσει αν οι αρχιτεκτονικές κάνουν overfit, αλλά και να συγκριθεί η απόδοση μεταξύ διαφορετικών αρχιτεκτονικών.

Συνεπώς, στο Σχήμα 4.6 παρουσιάζεται το Accuracy των MLP που προκύπτουν για διαφορετικές αρχιτεκτονικές ανά επανάληψη στο training και στο test set. Με διακεκομμένες γραμμές εμφανίζεται το Accuracy στο test-set, ενώ με τις συμπαγείς γραμμές το Accuracy στο training set. Να σημειωθεί πως η ετικέτα κάθε αρχιτεκτονικής αντιστοιχεί στην παράμετρο `hidden_layer_sizes` που αναγράφεται στον Πίνακα 3.4, δηλαδή η Αρχιτεκτονική 1 αφορά την αλληλουχία (10, 10), η Αρχιτεκτονική 2 την (30, 20, 15), κ.ο.κ. Όπως είναι φανερό, η Αρχιτεκτονική 2 παρουσιάζει την καλύτερη απόδοση κατά την αξιολόγηση στο test set, ιδιαίτερα από την 400^η επανάληψη κι έπειτα.

Αξίζει να σημειωθεί πως για όλες τις αρχιτεκτονικές που παρουσιάζονται σε αυτό το Σχήμα χρησιμοποιούνται οι ίδιες παράμετροι που αναγράφονται στον Πίνακα 4.2. Ακόμη, καθώς το Scikit-learn δεν παρέχει κάποιο τρόπο για την εξαγωγή αυτού του είδους της καμπύλης με αυτόματο τρόπο, όπως οι περισσότερες βιβλιοθήκες που εστιάζουν στην εκπαίδευση

νευρωνικών δικτύων, η εκπαίδευση χωρίζεται σε διαφορετικά στάδια και σε κάθε στάδιο υπολογίζονται οι μετρικές χειροκίνητα¹³.

4.2.4 Επιλογή Σημαντικότερων Χαρακτηριστικών

Έχοντας παρουσιάσει τα αποτελέσματα ταξινόμησης των αλγορίθμων, είναι ενδιαφέρον να δοθεί μία ερμηνεία στις επιλογές των αλγορίθμων, αφού, άλλωστε, αυτό αποτελεί κι ένα από τα πλεονεκτήματα των “κλασικών” αλγορίθμων μηχανικής μάθησης. Για τον λόγο αυτόν, στη συνέχεια παρουσιάζονται τα σημαντικότερα χαρακτηριστικά που προκύπτουν από το σύνολο δεδομένων και τους αλγορίθμους ταξινόμησης.

Αρχικά, εφαρμόζεται η στατιστική δοκιμασία χ^2 στο training set κι επιλέγονται μέσω αυτής τα 10 σημαντικότερα χαρακτηριστικά ανά είδος¹⁴, τα οποία καταγράφονται στον Πίνακα 4.8. Όπως γίνεται αντιληπτό, στο σύνολο των ειδών εμφανίζονται κάποια κοινά αντικείμενα, μεταξύ των οποίων: person, book, tie, chair, κλπ. Για να γίνει εύκολα αντιληπτό το ποια αντικείμενα είναι κοινά και από πόσους αλγορίθμους εντοπίζονται τα 10 σημαντικότερα χαρακτηριστικά, παρουσιάζεται αντίστοιχο ιστόγραμμα στο Σχήμα 4.7.

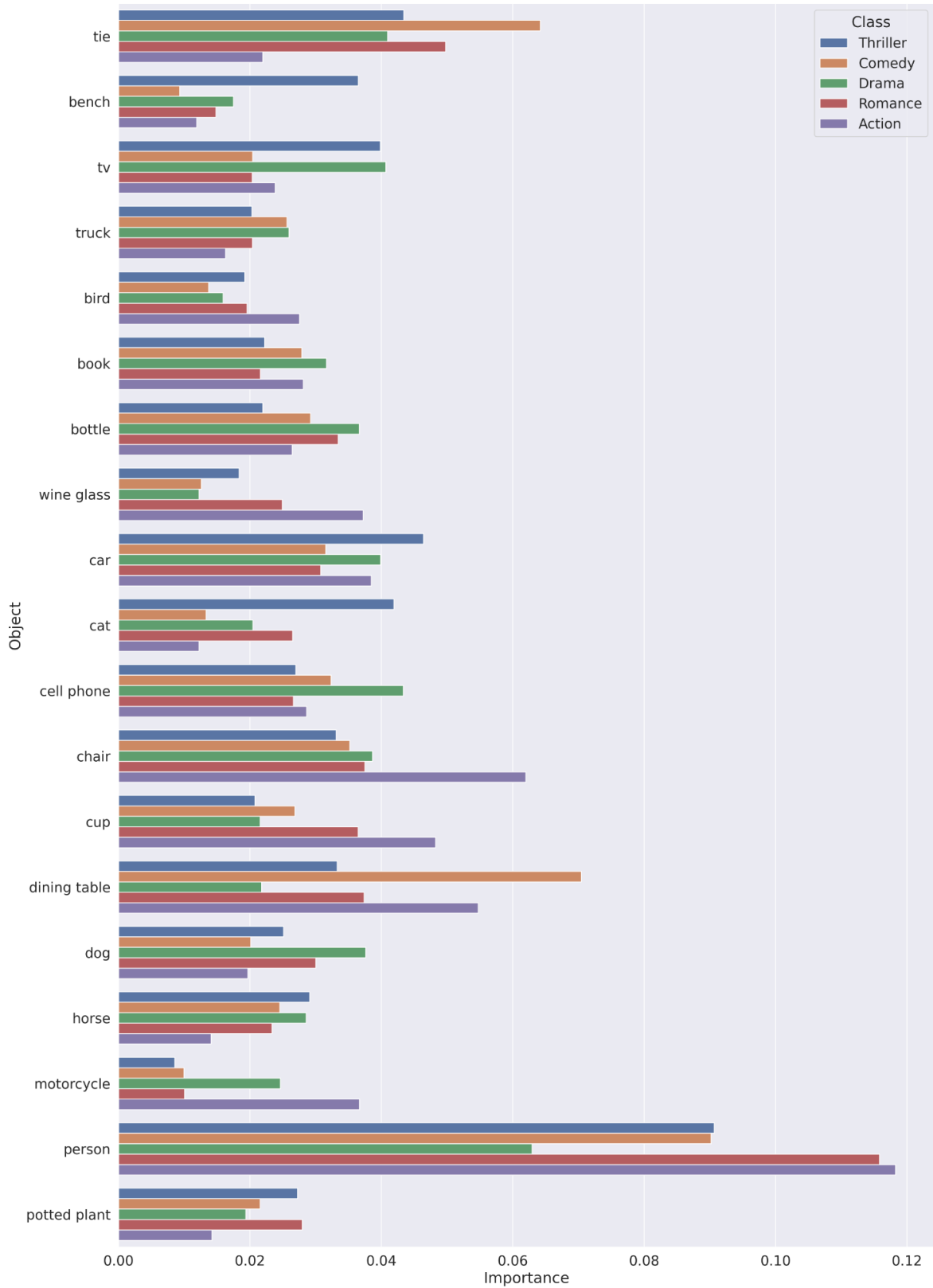
Αν και πολλά από τα αντικείμενα δεν μπορούν να χαρακτηριστούν ως αναμενόμενα για το αντίστοιχο είδος, μπορεί να γίνει η υπόθεση πως ίσως αυτά αποτελούν την ειδοποιό διαφορά για τη διάκριση των ειδών. Εκτός αυτού, κύριο ρόλο διαδραματίζει και η συχνότητα εμφάνισης του εκάστοτε αντικειμένου, η οποία από την εν λόγω τεχνική δεν είναι ξεκάθαρη. Πιο συγκεκριμένα, είναι πιθανό ένα αντικείμενο που έχει κριθεί σημαντικό από την τεχνική αυτή να μην εμφανίζεται συχνά στο trailer της ταινίας.

Πίνακας 4.8: Τα 10 σημαντικότερα χαρακτηριστικά ανά είδος ταινίας, σύμφωνα με τη δοκιμασία χ^2 στο training set.

| Drama | Comedy | Romance | Action | Thriller |
|--------------|---------------|----------------|---------------|-----------------|
| book | book | banana | Book | bench |
| bottle | bottle | bed | Bottle | book |
| cell phone | chair | bottle | Bowl | bowl |
| horse | cup | chair | Chair | car |
| laptop | horse | clock | Cup | chair |
| potted plant | person | couch | dining table | cup |
| sports ball | potted plant | horse | motorcycle | horse |
| stop sign | stop sign | person | person | person |

¹³ <https://stackoverflow.com/questions/46912557/is-it-possible-to-get-test-scores-for-each-iteration-of-mlpclassifier>

¹⁴ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html



Σχήμα 4.8: Ραβδόγραμμα που εμφανίζει τη σημασία του συνόλου των χαρακτηριστικών που προκύπτει ως η ένωση των 10 σημαντικότερων χαρακτηριστικών ανά είδος ταινίας, βάσει του Random Forest.

Με στόχο την εξακρίβωση των συχνοτήτων εμφάνισης κάθε αντικειμένου, εξάγονται ακόμα τα δέντρα αποφάσεων που παράγονται από τον αλγόριθμο Decision Tree για κάθε είδος ταινίας. Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί τις ίδιες παραμέτρους που επιτυγχάνουν τα υψηλότερα αποτελέσματα, με εξαίρεση το `max_depth`, το οποίο εδώ ισούται με 5, ώστε τα δέντρα να είναι πιο εύκολα ερμηνεύσιμα. Ακόμη, για να είναι πιο εύκολα συγκρίσιμες οι τιμές που εμφανίζονται σε κάθε δέντρο, πραγματοποιείται κανονικοποίηση με βάση το μέσο όρο και τη τυπική απόκλιση. Για την ακρίβεια, από κάθε τιμή του χαρακτηριστικού αφαιρείται ο μέσος όρος κι το υπόλοιπο διαιρείται με την τυπική απόκλιση. Με τις συγκεκριμένες παραμέτρους, ο αλγόριθμος επιτυγχάνει λίγο χαμηλότερα αποτελέσματα, με τη συνολική απόδοση να παρουσιάζεται στην ακόλουθη λίστα:

- Precision: 50%.
- Recall: 46%.
- F-Score: 45%.
- Exact Match Ratio: 0.2142.
- Hamming Loss: 0.2762.

Στο Σχήμα 4.9 παρουσιάζεται το δέντρο αποφάσεων για το είδος Drama. Από αυτό ενδιαφέρον παρουσιάζει το μονοπάτι στο δέντρο `laptop`, `tv`, `motorcycle`, `book`, `cat`, όπου αν ισχύει ότι εμφανίζονται σχετικά λίγες φορές μέσα στις σκηνές του trailer, η ταινία κατηγοριοποιείται ως Drama. Αν και τα περισσότερα χαρακτηριστικά δεν ερμηνεύονται εύκολα μεμονωμένα, το γεγονός ότι οι μοτοσυκλέτες δεν παίζουν σημαντικό ρόλο σε τέτοιου είδους ταινίες δεν αποτελεί έκπληξη.

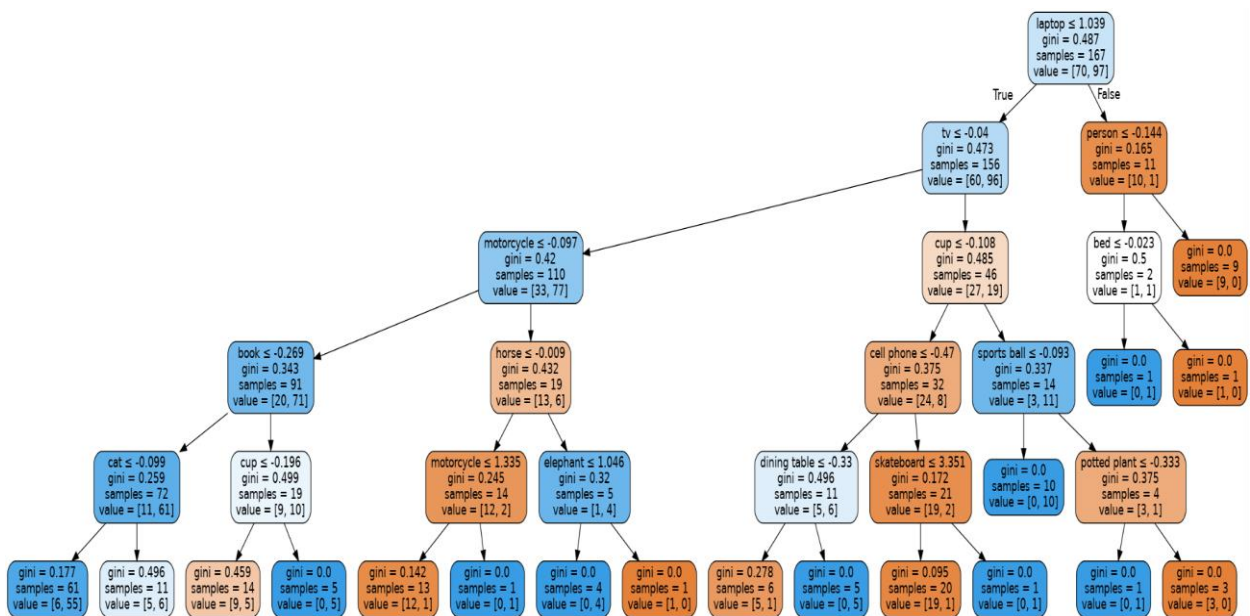
Στο Σχήμα 4.10 παρουσιάζεται το δέντρο απόφασης για το είδος Comedy. Εδώ το κυριότερο χαρακτηριστικό που χρησιμοποιείται για τον πρώτο διαχωρισμό των δεδομένων είναι το `dining table`. Για την ακρίβεια, αν εμφανίζονται λίγα `dining tables` στο trailer μίας ταινίας, τότε κατά πάσα πιθανότητα, η ταινία δεν ανήκει στο είδος Comedy, το οποίο είναι σχετικά λογικό, δεδομένου ότι είναι σχετικά σύνηθες σε τέτοιου είδους ταινίες να εντοπίζονται παρέες που συγκεντρώνονται γύρω από ένα τραπέζι.

Αναφορικά με το Σχήμα 4.11, στο οποίο παρουσιάζεται το δέντρο απόφασης για το είδος Romance, καθοριστικής σημασίας αποτελεί το χαρακτηριστικό `person`, αφού αν ανιχνεύονται σχετικά σπάνια άνθρωποι στο trailer της ταινίας, τότε κατά πάσα πιθανότητα η ταινία δεν ανήκει σε αυτό το είδος. Αυτό είναι λογικό, αφού αυτού του είδους οι ταινίες περιστρέφονται κυρίως γύρω από τις προσωπικές σχέσεις των πρωταγωνιστών, οπότε είναι αναμενόμενο να παίζουν κυρίαρχο ρόλο στο trailer της ταινίας.

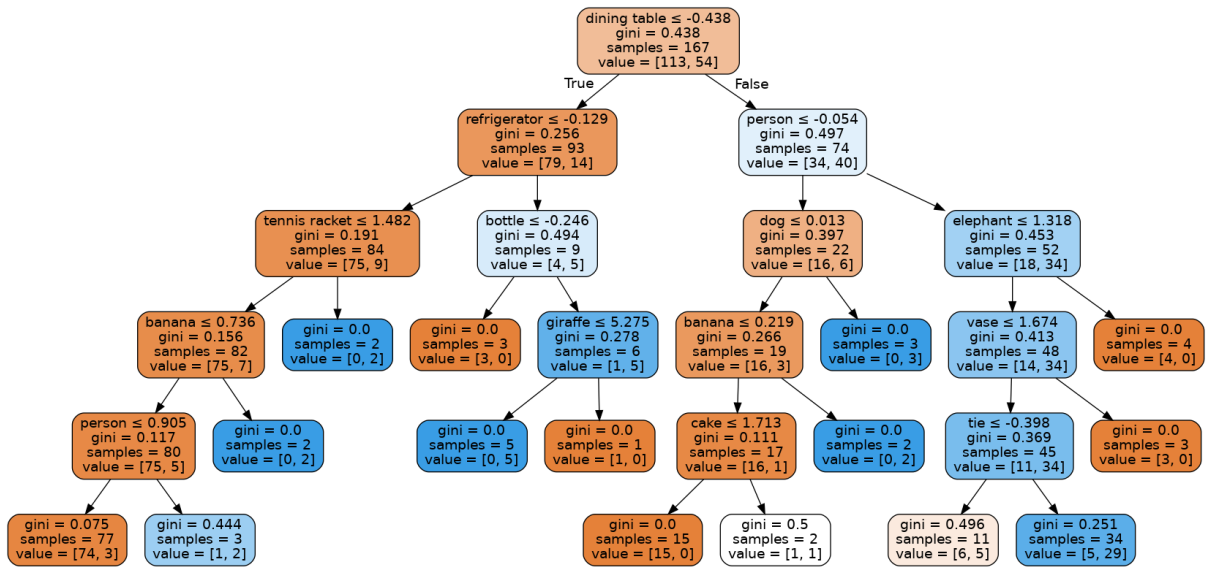
Στο Σχήμα 4.12 απεικονίζεται το δέντρο απόφασης για τις ταινίες που ανήκουν στο είδος Action. Σύμφωνα με αυτό, το πλήθος των σκηνών στο οποίο παρουσιάζονται αντικείμενα ως `person` και πάλι έχει υψηλή σημασία. Για την ακρίβεια, αν εμφανίζονται περισσότερα από το μέσο όρο των αντικειμένων `person` ανά `frame`, τότε με μεγάλη πιθανότητα η ταινία δεν ανήκει στο είδος Action. Από την άλλη, αν εμφανίζονται αρκετά άτομα, τότε θα πρέπει να εμφανίζονται αρκετά αμάξια ή πληκτρολόγια, ώστε η ταινία να ανήκει στο είδος αυτό. Οι παραπάνω κανόνες δεν είναι παράλογοι, αφού σε τέτοιου είδους trailer συχνά εμφανίζονται

εντυπωσιακές σκηνές στις οποίες δεν παίζουν τόσο έντονο ρόλο οι χαρακτήρες (π.χ. εκρήξεις, αγώνες δρόμου, κλπ), ενώ αντικείμενα όπως πληκτρολόγιο και αμάξια θα μπορούσαν να εμφανίζονται συχνά στα trailer τέτοιων ταινιών.

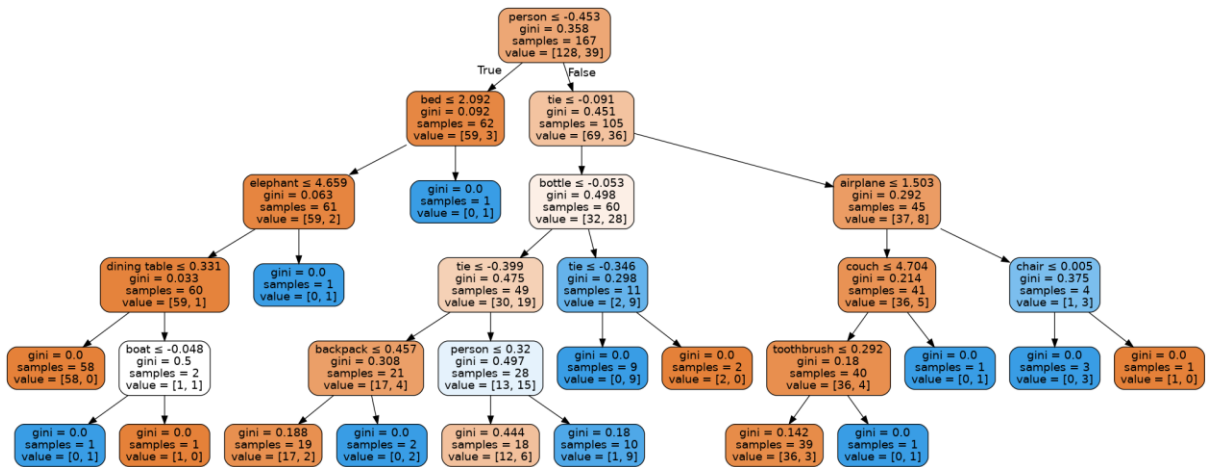
Τέλος, στο Σχήμα 4.13 παρουσιάζεται το δέντρο απόφασης για τις ταινίες Thriller. Παραδόξως, εδώ φαίνεται πως τα κυριότερα χαρακτηριστικά για το διαχωρισμό των ταινιών είναι το πλήθος των αντικειμένων person και cat. Για την ακρίβεια, σύμφωνα με το δέντρο, θα πρέπει να αναγνωρίζονται σχετικά λίγα αντικείμενα ως person αλλά παραπάνω από το μέσο όρο ως cat. Αν και μπορεί να υποστηρίξει κανείς πως σε τέτοιου είδους ταινίες, τα trailer θα προσπαθούσαν να εξάψουν την αγωνία του θεατή, οπότε δεν θα εμφανίζονταν συχνά άνθρωποι σε αυτά, δεν είναι ξεκάθαρο το γιατί παίζει τόσο υψηλό ρόλο το χαρακτηριστικό cat. Όμως, αξίζει να αναφερθεί πως ένα σημαντικό ποσοστό των δειγμάτων του training set που ανήκουν στο είδος Thriller δεν ταξινομούνται σωστά, οπότε η ορθότητα του δέντρου εδώ είναι αμφισβητήσιμη.



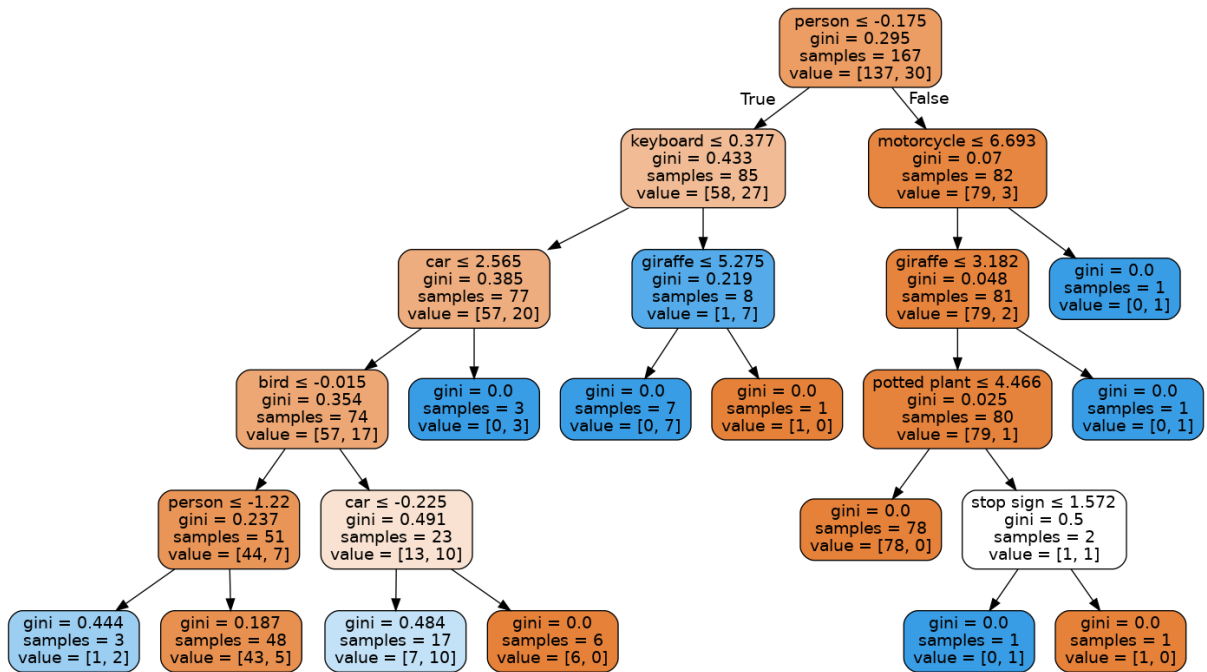
Σχήμα 4.9: Το δέντρο αποφάσεων για το είδος Drama.



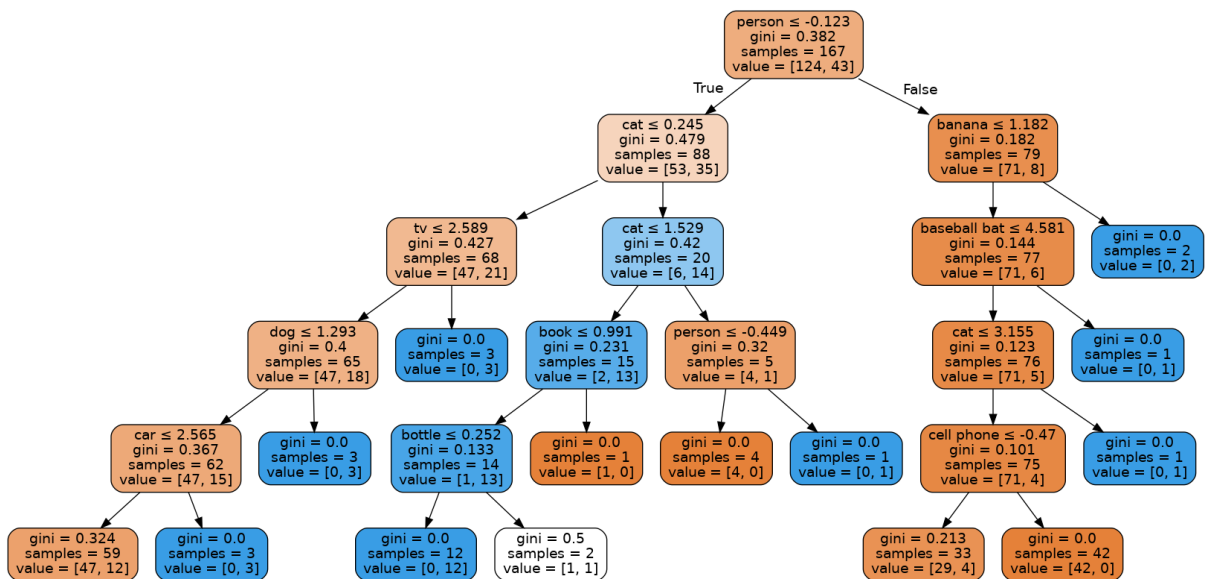
Σχήμα 4.10: Το δέντρο αποφάσεων για το είδος Comedy.



Σχήμα 4.11: Το δέντρο αποφάσεων για το είδος Romance.



Σχήμα 4.12: Το δέντρο αποφάσεων για το είδος Action.



Σχήμα 4.13: Το δέντρο αποφάσεων για το είδος Thriller.

4.3 Σύγκριση Αποτελεσμάτων

Με σκοπό την επαλήθευση της καταλληλότητας της χρησιμοποιούμενης μεθοδολογίας, πραγματοποιείται μια σύγκριση των εξαγόμενων αποτελεσμάτων. Ξεκινώντας από τα δεδομένα που διαδραματίζουν το πιο σημαντικό ρόλο στην αυτόματη αναγνώριση του είδους μιας ταινίας, τα αποτελέσματα της παρούσας εργασίας έρχονται σε συμφωνία με αυτά της βιβλιογραφίας. Επί παραδείγματι, και σύμφωνα με σχετική μελέτη, στις ταινίες Action

εντοπίζονται κυρίως άτομα (person) και αυτοκίνητα (car), ενώ στις ταινίες Romance παρατηρούνται και πάλι άτομα (Chu & Guo, 2017).

Συνεχίζοντας με την απόδοση των αλγορίθμων, στην έρευνα (Alrowaili, 2019), όπου ακολουθείται μία αρκετά όμοια διαδικασία με επίσης 5 διαφορετικά είδη ταινιών, επιτυγχάνεται μέσο precision 52% και μέσο recall 44.6% σε ένα μεγαλύτερο σύνολο δεδομένων (συνολικά 893 trailers). Η απόδοση αυτή επετεύχθη χρησιμοποιώντας, επίσης, τον αλγόριθμο Random Forest, αν και πραγματοποιήθηκαν πειράματα με νευρωνικό δίκτυο MLP, το οποίο όμως απέδωσε χειρότερα. Συγκριτικά, τα εξαγόμενα αποτελέσματα της παρούσας εργασίας είναι υψηλότερα, αφού επιτυγχάνεται 62% μέσο Precision και 56% μέσο Recall. Αξίζει να σημειωθεί πως η συγκεκριμένη δουλειά είναι μάλλον η πλησιέστερη της βιβλιογραφίας.

Στην έρευνα (Zhou et al., 2010) οι συγγραφείς εξάγουν διάφορα οπτικά χαρακτηριστικά του trailer μίας ταινίας, όπως το είδος κάθε σκηνής, αλλά όχι όμως τα αντικείμενα που συμπεριλαμβάνονται σε αυτή με στόχο την ταξινόμηση του είδους της ταινίας. Έτσι, κάνοντας ακόμη την παραδοχή πως μία πρόβλεψη είναι σωστή αν ο αλγόριθμος ταξινόμησης προβλέψει σωστά τουλάχιστον ένα από τα είδη της ταινίας, πετυχαίνουν αρκετά υψηλές αποδόσεις της τάξης του 74.7% Accuracy. Παρ' όλ' αυτά, και όπως έχει ήδη αναφερθεί, η συγκεκριμένη προσέγγιση έχει λάβει κριτική λόγω της συγκεκριμένης παραδοχής και τα αποτελέσματα δεν είναι ιδιαίτερα συγκρίσιμα, αφού στην παρούσα εργασία αντιμετωπίζεται το multi-label πρόβλημα ταξινόμησης χωρίς κάποια αντίστοιχη παραδοχή.

Τέλος, στην έρευνα (Simões et al., 2016) και όπως έχει ήδη αναφερθεί σε προηγούμενο κεφάλαιο, χρησιμοποιούνται CNNs για την εξαγωγή χαρακτηριστικών σε μορφή Bag of Words. Όμως, η συγκεκριμένη έρευνα χρησιμοποιεί ταινίες που ανήκουν σε 1 μόνο κατηγορία από τις 4 που επιλέγουν, κατατάσσοντας το πρόβλημα σε multi-class ταξινόμηση που είναι πιο απλό από τη multi-label ταξινόμηση. Ακόμη, χρησιμοποιούν και ακουστικά χαρακτηριστικά και συνολικά επιτυγχάνουν επιδόσεις της τάξης του 62% accuracy σε ένα αρκετά μεγαλύτερο σύνολο δεδομένων που αποτελείται από 1067 ταινίες.

5. Συμπεράσματα

Η παρακολούθηση ταινιών αποτελεί ένα τρόπο ψυχαγωγίας για μεγάλη μερίδα ανθρώπων, καθώς με αυτόν τον τρόπο προτιμούν να αξιοποιούν τον ελεύθερο χρόνο τους είτε ατομικά είτε σε ομάδες. Κάθε ταινία, ανάλογα με το είδος της πυροδοτεί μια διαφορετική αλληλουχία συναισθημάτων, που πιθανώς να μην είχε προκληθεί, αν ο χρόνος αυτός είχε δαπανηθεί διαφορετικά.

Η παρακολούθηση μιας ταινίας μπορεί να μεταβάλλει τη συναισθηματική κατάσταση του θεατή και να τον κάνει να ταυτιστεί με τους πρωταγωνιστές και την ιστορία που διαδραματίζεται, βοηθώντας τον να ξεφύγει από τα προβλήματα της καθημερινότητάς του. Πολλές φορές, και ανάλογα με το είδος της, μια ταινία μπορεί να διεγείρει το ενδιαφέρον του θεατή για τα γεγονότα που εξελίσσονται στην οθόνη του και να προκαλέσει μια σειρά ερωτημάτων, ιδιαίτερα αν τα γεγονότα αυτά αφορούν σε κάποιο ιστορικό γεγονός ή τεχνολογικό επίτευγμα.

Εκτός, όμως, από τις προαναφερόμενες ωφέλειες, η παρακολούθηση μιας ταινίας μπορεί να αποτελέσει και κοινωνική συναναστροφή εφόσον πραγματοποιείται με την παρουσία περισσότερων ατόμων. Συνεπώς, η ανάπτυξη ανθρώπινων σχέσεων και αλληλεπιδράσεων ευδοκιμεί, προάγοντας την επικοινωνία και τον διάλογο μέσω διαφόρων σχολίων που αφορούν την πλοκή και το σενάριο της ταινίας.

Η παραγωγή μιας κινηματογραφικής ταινίας αποτελεί μια απαιτητική, κοστοβόρα και χρονοβόρα διαδικασία, γι' αυτό και οι εμπειρογνώμονες του κλάδου αφιερώνουν πολλούς πόρους, προκειμένου να κατανοήσουν την προτίμηση του κοινού και να προσελκύσουν όσο το δυνατόν περισσότερους θεατές. Ωστόσο, ένας παράγοντας υψίστης σημασίας για την επιτυχία μιας ταινίας είναι το είδος αυτής.

Το είδος μιας ταινίας εξυπηρετεί στην κατηγοριοποίηση αυτής με τρόπο τέτοιον, ώστε να προϋδεάσει το θεατή ως προς το γενικότερο πλαίσιο και θεματολογία της ταινίας. Η πρόωθηση κάθε ταινίας λαμβάνει χώρα, μεταξύ άλλων, με τη βοήθεια ενός trailer, το οποίο αποτελεί ένα ολιγόλεπτο βίντεο με διάφορες σκηνές και διαλόγους των ηθοποιών, το οποίο προϋδεάζει τον θεατή αναφορικά με το τι αναμένεται να παρακολουθήσει.

Προκύπτει, λοιπόν, εύλογα πως τόσο το είδος, όσο και το trailer μιας ταινίας παίζουν σημαντικό ρόλο στην απήχηση αυτής και μπορούν να λειτουργήσουν συνδυαστικά, ώστε να προσελκύσουν το κοινό. Παρά ταύτα, είναι πιθανό κατά την σύλληψη και την παραγωγή μιας ταινίας, να μην έχει διευκρινιστεί με σαφήνεια το είδος της, ενώ παρατηρείται συχνά πως μια ταινία συνδυάζει πολυάριθμα και ποικίλα στοιχεία από διαφορετικά είδη ταινιών.

Κατά συνέπεια, η σαφής κατάταξη μιας ταινίας σε είδος μπορεί να είναι δύσκολη, ενώ η λανθασμένη κατηγοριοποίηση μπορεί να αποφέρει αρνητικά αποτελέσματα στην απήχηση της. Λαμβάνοντας υπόψη πως το μεγαλύτερο ποσοστό των ατόμων πριν προβούν στην

παρακολούθηση μιας ταινίας, παρακολουθούν το trailer της, η αυτόματη αναγνώριση του είδους μιας ταινίας από το trailer μπορεί να είναι εξαιρετικά ωφέλιμη.

Η αυτοματοποίηση αυτή μπορεί κυρίως να εξοικονομήσει αρκετό χρόνο και πόρους, μειώνοντας πιθανά λάθη, αυξάνοντας την επιτυχία μιας ταινίας και δίνοντας το έναυσμα για παραγωγή παρόμοιων ταινιών. Επιπρόσθετα, ένα τέτοιο εγχείρημα μπορεί να γίνει πραγματικότητα με τη χρήση τεχνικών και αλγορίθμων της TN.

Κατ' επέκτασιν, η παρούσα εργασία επικεντρώνεται στην αυτόματη αναγνώριση του είδους μιας ταινίας, μέσω των trailers. Πιο συγκεκριμένα, και στα πλαίσια της εργασίας, κατασκευάστηκε ένα σύνολο δεδομένων από 209 trailers με τη βοήθεια του IMDb και του YouTube. Τα trailers αυτά, αφού έλαβαν την απαραίτητη προεπεξεργασία, χρησιμοποιήθηκαν για την εφαρμογή της μεθόδου ανίχνευσης αντικειμένων YOLO.

Έτσι, κατασκευάστηκε ένα σύνολο δεδομένων ταξινόμησης κι έγινε φανερό πως τα αντικείμενα που πρωταγωνιστούν σε κάθε είδος διαφέρουν σε κάποιον σημαντικό βαθμό. Πιο αναλυτικά, φάνηκε πως το πόσο συχνά εμφανίζονται άτομα στις σκηνές ενός trailer παίζει καθοριστικό ρόλο στα περισσότερα είδη ταινιών, αλλά και πως περισσότερο συγκεκριμένα χαρακτηριστικά, όπως τα αυτοκίνητα, καθορίζουν σε μεγάλο βαθμό συγκεκριμένα είδη ταινιών.

Ως προς την ταξινόμηση των ειδών ταινιών, έγινε χρήση μιας ποικιλίας αλγορίθμων, με τα αποτελέσματα αυτών να κυμαίνονται σε ικανοποιητικά επίπεδα. Την υψηλότερη απόδοση φάνηκε να καταλαμβάνει ο αλγόριθμος Random Forest με 62% μέσο Precision. Παρ' όλ' αυτά, έγινε αντιληπτό πως η αυτόματη αναγνώριση με παραδοσιακούς αλγορίθμους μηχανικής μάθησης δεν αποφέρει αρκετά υψηλά αποτελέσματα, επιβεβαιώνοντας την πολυπλοκότητα του προβλήματος και φανερώνοντας πως μόνο η χρήση κάποιας μεθόδου ανίχνευσης αντικειμένων δεν επαρκεί.

Συνεπώς, υπάρχει σημαντικό περιθώριο βελτίωσης στη διαδικασία ταξινόμησης, αλλά ιδιαίτερα και στη διαδικασία εξαγωγής χαρακτηριστικών, αφού είναι πιθανό σε αυτήν να οφείλονται κάποια αναπάντεχα συμπεράσματα σχετικά με τα κυριότερα χαρακτηριστικά. Φυσικά, κάποια βελτίωση θα μπορούσε να παρατηρηθεί αν υπήρχε διαθέσιμη περισσότερη υπολογιστική ισχύς, αφού στη διαδικασία προεπεξεργασίας τα trailers συρρικνώθηκαν σε μόλις 128x128 pixels, κάτι που δε βοηθάει ιδιαίτερα στην ακριβή ανίχνευση αντικειμένων. Ακόμη, το σύνολο δεδομένων που κατασκευάστηκε είναι σχετικά μικρό, γεγονός που επίσης δικαιολογεί εν μέρει κάποιες ανακρίβειες κατά την ταξινόμηση.

Εν κατακλείδι, για την επίτευξη υψηλότερων αποδόσεων και ως μελλοντική βελτίωση μπορεί να ακολουθηθεί μια μεθοδολογία, η οποία να περιλαμβάνει αλγορίθμους βαθιάς μάθησης και να συνδυάζει πολλά και διαφορετικά χαρακτηριστικά που μπορούν να εξαχθούν από trailers. Για παράδειγμα, ενδιαφέρον θα παρουσίαζε η διερεύνηση του κατά πόσο παίζει σημαντικό ρόλο η χρονική σειρά με την οποία εμφανίζονται τα αντικείμενα στο trailer μίας ταινίας, μιας και η προσέγγιση της παρούσας εργασίας λαμβάνει υπόψη μόνο τη συνολική συχνότητα

εμφάνισης ενός αντικειμένου. Για την επίτευξη αυτού, θα χρειαζόταν μία πιο περίπλοκη αρχιτεκτονική που κατά πάσα πιθανότητα θα αξιοποιούσε κάποιο είδος DNN, τα αποτελέσματα του οποίου, όμως, δεν θα ήταν εύκολα ερμηνεύσιμα.

Βιβλιογραφία

- Κοντονή, Β. (2015). Η ιστορία και η χρηματοδότηση του ευρωπαϊκού κινηματογράφου: η ελληνική περίπτωση. *Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών*.
- Σκοπετέας, Ι. (2015). *Η δημιουργία της μυθοπλαστικής αφήγησης και τα είδη των κινηματογραφικών ταινιών*. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. <http://hdl.handle.net/11419/5729>
- Τάγκα, Ε. Κ. (2014). *Τεχνικές κινηματογραφικής και τηλεοπτικής παραγωγής—Από τις υπερπαραγωγές στις χαμηλού προϋπολογισμού ταινίες*. Aristotle University of Thessaloniki.
- Alrowaili, R. (2019, December). Movie Genre Recognition. *Seidenberg School of Computer Science, Pace University*.
- Arijon, D. (1991). *Grammar of the film language*. Silman-James Press.
- Banham, M., & Brandon, J. R. (1995). *The Cambridge guide to theatre*. Cambridge University Press.
- Chu, W. T., & Guo, H. J. (2017, October). Movie genre classification based on poster images with deep neural networks. *In Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes*, pp. 39-45.
- Ekvall, S., Kragic, D., & Jensfelt, P. (2007). Object detection and mapping for service robot tasks. *Robotica*, 25(2), pp. 175-187.
- Fei, N., & Zhang, Y. (2019, December). Movie genre classification using TF-IDF and SVM. *In Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City*, (pp. 131-136).
- García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer International Publishing.

- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), pp. 1.
- Huang, Q., Xiong, Y., Rao, A., Wang, J., & Lin, D. (n.d.). Movienet: A holistic dataset for movie understanding. *arXiv preprint*.
- Huang, Y. F., & Wang, S. H. (2012, December). Movie genre classification using svm with audio and video features. *In International Conference on Active Media Technology. Springer.*, pp. 1-10.
- Jain, S. K., & Jadon, R. S. (2009, September). Movies genres classifier using neural network. *In 2009 24th International Symposium on Computer and Information Sciences. IEEE.*, pp. 575-580.
- Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends in Computer Graphics and Vision.*, 12(1-3), pp. 1-308.
- Joo, J. S. (2020). Netflix and Changes in the Hollywood Film Industry. *Journal of Industrial Convergence*, 18(5), pp. 36-41.
- Karasulu, B., & Korukoglu, S. (2013). Moving object detection and tracking in videos. *In Performance Evaluation Software. Springer*, pp. 7-30.
- Kastrinaki, V., Zervakis, M., & Kalaitzakis, K. (2003). A survey of video processing techniques for traffic applications. *Image and vision computing*, 21(4), pp. 359-381.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, L. C. (2014). Microsoft coco: Common objects in context. *European conference on computer vision*, pp. 740-755.

- Liu, L., Ouyang, W., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, 128(2), pp. 261-318.
- McKee, R. (1999). *Story: Substance, Structure, Style and the Principles of Screenwriting*. Methuen Publishing Ltd.
- Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzian, A., Duffy, N., & Hodjat, B. (2019). Evolving deep neural networks. In *Artificial intelligence in the age of neural networks and brain computing* (pp. 293-312). Academic Press.
- Nabi, R. L., & Clark, S. (2008). Exploring the limits of social cognitive theory: Why negatively reinforced behaviors on TV may be modeled anyway. *Journal of Communication*, 58(3), pp. 407-427.
- Neale, S. (1990). Questions of genre. *Screen*, 31(1).
- New York Film Academy. (2016, August 6). *12 of the Most Popular Movie Genres By Country*. New York Film Academy. Retrieved April, 2021, from <https://www.nyfa.edu/student-resources/12-of-the-most-popular-movie-genres-by-country/>
- Ongsulee, P. (2017). Artificial intelligence, machine learning and deep learning. In *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*. *IEEE.*, pp. 1-6.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 12, pp. 2825-2830.

- Rasheed, Z., & Shah, M. (2002). Movie genre classification by exploiting audio-visual features of previews. *Object Recognition Supported by User Interaction for Service Robots*, 2(2), pp 1086-1089. <https://doi.org/10.1109/ICPR.2002.1048494>
- Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Schatz, T. (2013). Conglomerate Hollywood and American independent film. *American Independent Cinema: Indie, Indiewood and Beyond*, pp. 127-139.
- Shambharkar, P. G., & Doja, M. N. (2020). Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences. *Multimedia Tools and Applications*, 79(29), pp. 21197-21222.
- Shambharkar, P. G., Thakur, P., Imadoddin, S., Chauhan, S., & Doja, M. N. (2020). Genre Classification of Movie Trailers using 3D Convolutional Neural Networks. *In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 850-858.
- Simões, G. S., Wehrmann, J., Barros, R. C., & Ruiz, D. D. (2016, July). Movie genre classification with convolutional neural networks. *In 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 259-266.
- Singh, J., Goyal, G., & Gupta, S. (2019). FADU-EV an automated framework for pre-release emotive analysis of theatrical trailers. *Multimedia Tools and Applications*, 78(6), pp. 7207-7224.
- Singhal, A., Kumar, P., Saini, R., Roy, P. P., Dogra, D. P., & Kim, B. G. (2018). Summarization of videos by analyzing affective state of the user through crowdsourcing. *Cognitive Systems Research*, 52, pp. 917-930.
- Sorower, M. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, 18, pp. 1-25.

- Wehrmann, J., & Barros, R. C. (2017). Movie genre classification: A multi-label approach based on convolutions through time. *Applied Soft Computing*, *61*, pp. 973-982.
- Wehrmann, J., Simões, G. S., Barros, R. C., & Cavalcante, V. F. (2018). Adult content detection in videos with convolutional and recurrent neural networks. *Neurocomputing*, *272*, pp. 432-438.
- Yadav, A., & Vishwakarma, D. K. (2020). A unified framework of deep networks for genre classification using movie trailer. *Applied Soft Computing*, *96*, pp. 106624.
- Zhou, H., Hermans, T., Karandikar, A. V., & Rehg, J. M. (2010, October). Movie genre classification via scene categorization. *In Proceedings of the 18th ACM international conference on Multimedia*, pp. 747-750.
- Zou, X. (2019, August). A Review of Object Detection Techniques. *In 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*. IEEE., pp. 251-254.
- Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*.