ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
SCHOOL OF BUSINESS
ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΟΙΚΗΤΙΚΗ ΕΠΙΣΤΗΜΗ & ΤΕΧΝΟΛΟΓΙΑ
MSc IN MANAGEMENT SCIENCE & TECHNOLOGY

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΓΕΩΡΓΙΑΣ ΚΡΙΜΠΕΝΗ

ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ ΝΙΚΗΤΩΝ ΙΠΠΟΔΡΟΜΙΑΣ

ΠΡΑΚΤΙΚΗ ΆΣΚΗΣΗ ΣΤΟΝ ΟΜΙΛΟ ΟΠΑΠ

Επιβλέπων : Αγγελική Πουλυμενάκου – Διδάσκων ΟΠΑ

Υποβληθείσα ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος (MSc) στη Διοικητική Επιστήμη και Τεχνολογία

Αθήνα, Δεκέμβριος 2019

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ και ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΓΕΩΡΓΙΑΣ ΚΡΙΜΠΕΝΗ

ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ ΝΙΚΗΤΩΝ ΙΠΠΟΔΡΟΜΙΑΣ

ΠΡΑΚΤΙΚΗ ΆΣΚΗΣΗ ΣΤΟΝ ΟΜΙΛΟ ΟΠΑΠ

Επιβλέπων : Αγγελική Πουλυμενάκου – Διδάσκων ΟΠΑ

Υποβληθείσα ως μέρος των απαιτήσεων για την απόκτηση

Μεταπτυχιακού Διπλώματος (MSc) στη Διοικητική Επιστήμη και Τεχνολογία

Αθήνα, Δεκέμβριος 2019

Βεβαίωση εκπόνησης Διπλωματικής εργασίας

«Δηλώνω υπεύθυνα ότι η συγκεκριμένη μεταπτυχιακή εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών του ΠΜΣ στη Διοικητική Επιστήμη και Τεχνολογία του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του Οικονομικού Πανεπιστημίου Αθηνών έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών στην Ελλάδα ή το εξωτερικό. Η εργασία αυτή έχοντας εκπονηθεί από εμένα, αντιπροσωπεύει τις προσωπικές μου απόψεις επί του θέματος. Οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης διπλωματικής αναφέρονται στο σύνολό τους, δίνοντας πλήρεις αναφορές στους συγγραφείς, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο».

(Υπογραφή)

.....

ΚΡΙΜΠΕΝΗ ΓΕΩΡΓΙΑ

Φοιτητής MSc στη Διοικητική Επιστήμη και Τεχνολογία



Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη μεθοδολογίας για την καταγραφή και πρόβλεψη των αποτελεσμάτων των κουρσών της κάθε ιπποδρομιακής συγκέντρωσης με βάση τα δεδομένα που παρέχει ο Ελληνικός Ιππόδρομος στους φιλίππους (πχ επίσημο πρόγραμμα ιπποδρομιών, εβδομαδιαία νέα, προγνωστικά, αποτελέσματα των ιπποδρομιών). Για την συγκέντρωση των δεδομένων χρησιμοποιήθηκε η κεντρική ιστοσελίδα του ιπποδρόμου. Η αναγνώριση και η καταγραφή των δεδομένων έγινε με την χρήση του Microsoft Excel και στη συνέχεια γίνεται μια πρόταση για καλύτερη αξιοποίηση των δεδομένων ,με τη βοήθεια των προγραμμάτων SAS Enterprise Guide, Rapidminer και R Studio, προκειμένου να γίνονται οι προβλέψεις με μεγαλύτερη ακρίβεια.

Παράλληλα, στη πρακτική εργασία τα κύρια καθήκοντα ήταν η ανάλυση δεδομένων (Data Analysis) καθώς και η υποστήριξη διαχείρισης εσωτερικού και έργων (Intern & Project Management Support). Ο στόχος ήταν η βελτίωση της εργασίας των εργαζομένων καθώς επίσης και η δημιουργία του πρώτου βιβλίου που θα κατέγραφε τα ιστορικά στοιχεία των ελληνικών ιπποδρομιών (FactBook) του 2019.

Συγκεκριμένα, έγινε μελέτη των συστημάτων και των μοντέλων πρόβλεψης που υπάρχουν στο εξωτερικό. Στη συνέχεια, έγινε καταγραφή όλων των δεδομένων του έτους 2019 των ελληνικών ιπποδρομίων και δημιουργία ενός ελληνικού μοντέλου πρόβλεψης ως προς τα δεδομένα που υπήρχαν.

Η μεθοδολογία αυτή μπορεί να γίνει οδηγός για την πραγματοποίηση εξωτερικών ή εσωτερικών μετρήσεων πρόβλεψής στις ελληνικές ιπποδρομίες, με ελάχιστες αλλαγές ως προς τις μεταβλητές των μετρήσεων. Η γενικότητα της μεθοδολογίας έγκειται στο γεγονός ότι έχουν καταγραφεί όλα τα βήματα, διεξαγωγής των σωστών δεδομένων που συμβάλλονται προκειμένου να υπάρχουν όλα τα στάδια διεξαγωγής των μετρήσεων και των προβλέψεων.

Λέξεις Κλειδιά: Ιππόδρομος, Ελληνικός Ιππόδρομος, αναλύσεις δεδομένων, υποστήριξη διαχείρισης έργων, προβλέψεις, ΟΠΑΠ, Markopoulos Park, SAS, RapidMiner, R-studio.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Abstract

The purpose of this dissertation is to develop a methodology for recording and predicting the results of each horse-race based on the data provided by the Hellenic Hippodrome to Philippi (e.g. official horse racing program, weekly news, horse racing results). The horse race's central website was used to detect the data. Data were identified and recorded using Microsoft Excel and then a proposal was made to make better use of the data with the help of SAS Enterprise Guide, RapidMiner and R Studio programs so that even more accurate predictions are made.

At the same time, the main tasks as an intern were data analysis (Data Analysis), as well as internal and project management support (Intern & Project Management Support). The aim was to improve the work of the employees and moreover create the first book to record the historical clues of the Greek horseracing (FactBook) of 2019.
Specifically, a study was conducted for the forecasting systems and models that exist abroad.

Specifically, the systems and models available abroad were studied. Subsequently, all the figures for the year 2019 of the Greek horse racing were recorded and an application of the Greek model was created based on the data that existed.

This methodology can be used as a guide for making external or internal forecasting measurements in Greek horseracing, with little change in the measurement variables.
The generality of the methodology lies in the fact that all the steps involved in performing the correct data are recorded in order to have all stages of the measurements and forecasts carried out.

Keywords: Horse Race; Hellas Horse Races; Data Analysis; Project Management Support; forecast; OPAP; Markopoulos Park; SAS; RapidMiner; R-studio.

Περιεχόμενα

1 Εισαγωγή 6

1.1 Η Ιστορία των Ιπποδρομιών και η φίλιππη κοινότητα 6

1.2 Αντικείμενο διπλωματικής 6

1.2.1 Συνεισφορά 7

1.3 Δομή της εργασίας 8

2 Το πρόβλημα της πρόβλεψης αθλητικών γεγονότων 9

2.1 Deep Learning 10

2.2 Εφαρμογή στις προβλέψεις αγώνων 10

3 Θεωρητικό υπόβαθρο 11

3.1 Γενικευμένο γραμμικό μοντέλο 11

3.2 Μέθοδος του τυχαίου δάσους 13

4 Προετοιμασία και ανάλυση των ιστορικών δεδομένων ιπποδρομιακών αγώνων 17

4.1 Διαθέσιμα ιστορικά στοιχεία 18

4.2 Διαλογή των δεδομένων 20

5 Συγκέντρωση και επιλογή δεδομένων 22

5.1 Συμψηφισμός δεδομένων 22

5.2 Υπολογισμός των ανεξάρτητων μεταβλητών 23

5.3 Υλοποίηση του μοντέλου πρόβλεψης με κώδικα 26

6 Αξιολόγηση του μοντέλου πρόβλεψης 28

6.1 Παράμετροι αξιολόγησης 28

6.2 Σύστημα αξιολόγησης 28

6.3 Οργάνωση του συστήματος αξιολόγησης 29

6.4 Αποτελέσματα 30

6.5 Σύνοψη συμπερασμάτων αξιολόγησης 31

7 Τεχνικές λεπτομέρειες 32

7.1 Λεπτομέρειες υλοποίησης 32

7.1.1 Πίνακας συσχέτισης 32

7.1.2 Κώδικας r 32

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

7.2 Πλατφόρμες και προγραμματιστικά εργαλεία 34

8 Κύρια Στοιχεία της Πρακτικής Άσκησης 35

8.1 Τομέας : Περιγραφή του τμήματος (ή τμήματα) στο οποίο εργαστήκατε και ποια τα καθήκοντά σας. 35

8.2 Περιγραφή της θέσης σας στο τμήμα (ή τμήματα) 36

9 Έργα και Δραστηριότητες 37

9.1 Έργα, Δραστηριότητες και Προγραμματισμός 37

9.2 Ανάλυση Κύριων Έργων/Δραστηριοτήτων 37

9.3 Παραδοτέα του έργου 38

10 Γνώση, Δεξιότητες και Κριτική αξιολόγηση 39

10.1 Γνώσεις και Δεξιότητες 39

10.2 Κριτική και Προτάσεις 39

11 Επίλογος 40

11.1 Σύνοψη και συμπεράσματα 40

11.2 Μελλοντικές επεκτάσεις 40

12 Βιβλιογραφία 41

13 Παράρτημα 43

13.1 Κώδικας r 43

13.2 Πηγές δεδομένων [πηγή: https://www.markopoulopark.gr/] 50

13.2.1 Επίσημο Πρόγραμμα Ιπποδρομιών 50

13.2.2 Αποτελέσματα (ιπποδρομιακών συγκεντρώσεων ανά κούρσα) 51

13.2.3 Rating ίππων 52

13.2.4 Εβδομαδιαία νέα ιπποδρομιών 54

13.2.5 Factbook 2018 55


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1 *Εισαγωγή*

1.1 Η Ιστορία των Ιπποδρομιών και η φίλιππη κοινότητα

Η ιπποδρομία είναι ένας αγώνας ταχύτητας μεταξύ ίππων (αλόγων) από την αρχαιότητα έως και σήμερα. Ο νικητής ίππος στο παρελθόν ήταν αυτός που τερμάτιζε πρώτος, ενώ σήμερα δεν αρκεί να τερματίσει πρώτός αλλά να βγουν αρνητικά και τα αποτελέσματα doping . Ο έλεγχος doping γίνεται για να αποδειχθεί πως ο νικητής ίππος δεν έκανε χρήση απαγορευμένων ουσιών.

Στους Ολυμπιακούς αγώνες της Ολυμπίας το 648 π.Χ. διεξήχθη η ιπποδρομία «ίππου κέλητος» (οι ίπποι έτρεχαν έξι φορές φορές τον ίδιο δρόμο) , το 556 π.Χ. έγινε «πώλου κέλητος» (τέσσερις φορές το δρόμο), και το 496πΧ «κάλπης» όπου έτρεχαν φοράδες (δεν αναγράφεται πόσες φορές έτρεχαν στο δρόμο, όμως αναφέρεται πως ο αναβάτης κατέβαινε από την φοράδα στη τελευταία στροφή και έτρεχε μαζί της έως το τέρμα. Αυτός ο αγώνας καταργήθηκε λίγα χρόνια μετά το 444 π.Χ.)

Στη σημερινή εποχή οι ιπποδρομιακές συγκεντρώσεις χωρίζονται σε τρείς κατηγορίες κουρσών με καθαρόαιμους ίππους: η πρώτη είναι με αναβάτη, η δεύτερη είναι με τροχασμό (δηλαδή οι ίπποι σέρνουν αμαξίδιο με αμαξηλάτη) και η τρίτη κατηγορία είναι μετ' εμποδίων (γίνονται άλματα κατά τη διάρκεια της κούρσας).

Στο παρελθόν η ιπποτροφία ήταν γνώρισμα ευγενικής καταγωγής και κοινωνικής τάξης. Αυτή η θεωρία εξακολουθεί να υπάρχει έως και σήμερα, μόνο που έχει αλλοιωθεί λίγο κυρίως στη πρώιμη εποχή και στις υποανάπτυκτες χώρες. Για παράδειγμα, οι κούρσες στον χώρο τον ιπποδρομιών δεν γίνονται για το «ευ αγωνίζεσθαι» αλλά για το έπαθλο και κατ' επέκταση τα χρηματικά κέρδη όχι μόνο του ιδιοκτήτη, προπονητή, αναβάτη αλλά και για τους φιλίππους – δηλαδή τους οπαδούς - των ιπποδρομιακών συγκεντρώσεων.

1.2 Αντικείμενο διπλωματικής

Αντικείμενο της παρούσας διπλωματικής είναι η καταγραφή των στοιχείων και των παραμέτρων που χρησιμοποιούνται για την δημιουργία μοντέλου πρόβλεψης του νικητή ίππου κάθε ιπποδρομίας. Ακόμη γίνεται αναφορά στα καθήκοντα μου ανατέθηκαν και τις δραστηριότητες στις οποίες συμμετείχα κατά τη διάρκεια της πρακτικής μου άσκησης στις Ελληνικές Ιπποδρομίες στο Μαρκόπουλο – μια θυγατρική εταιρεία του ΟΠΑΠ. Η δημοφιλία των ιπποδρομιών είναι αποτέλεσμα της μεγάλης αύξησης της φίλιππης κοινότητας, η οποία βρίσκει διασκέδαση παρακολουθώντας ιπποδρομίες. Η επαγγελματική ενασχόληση αναβατών και προπονητών με την εκτροφή και την εκγύμναση ίππων αποκλειστικά για τις ιπποδρομίες έχει συμβάλλει στην ανάπτυξη ενός μοντέλου υπολογισμού της πιθανότητας νίκης κάθε συμμετέχοντος στο γεγονός της

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ιπποδρομίας πριν τη διεξαγωγή της . Ένα πολύ μεγάλο κύκλωμα που περιλαμβάνει στοιχηματικές εταιρείες , εκτροφείς , χώρους διεξαγωγής ιπποδρομιών και -φυσικά – το κοινό έχουν δημιουργήσει μια βιομηχανία θεάματος που , όπως αναμενόταν, δεν άφησε ασυγκίνητη ούτε την επιστημονική κοινότητα .

Από επιστημονικής πλευράς, η εκτίμηση της πιθανότητας νίκης των συμμετεχόντων σε ανταγωνιστικά γεγονότα, όπως είναι όλα τα αθλητικά δρώμενα, αποτελεί μια μεγάλη πρόκληση για τα κλασσικά μοντέλα πρόβλεψης σαν την παλινδρόμηση (regression) και την ταξινόμηση (classification) . Τα παραπάνω δεν είναι σχεδιασμένα με στόχο την ενσωμάτωση του ανταγωνιστικού στοιχείου, οπότε η πιθανότητα επιτυχίας ενός συμμετέχοντα εξαρτάται όχι μόνο από την ατομική του ικανότητα αλλά και από αυτή των ανταγωνιστών του. Ο κύριος στόχος της εργασίας είναι, πρώτον, η κριτική αξιολόγηση των υπαρχόντων εργαλείων πρόβλεψης που χρησιμοποιούνται και -δεύτερον- ο προσδιορισμός της βέλτιστης μεθόδου συλλογής στοιχείων που αφορούν την κούρσα αλλά και τους ίδιους τους συμμετέχοντες και αξιοποιούνται ώστε να δίνουν ακριβή αποτελέσματα μαζί με τη χρήση υπολογιστικών προγραμμάτων για την εκτίμηση της τελικής κατάταξης των ίππων προτού αυτοί αγωνιστούν σε ιπποδρομία. Έπειτα θα δοθεί βάση σε συστήματα αξιολόγησης των μοντέλων με βάσει τα αποτελέσματα που αυτά προτείνουν – σε συνάρτηση πάντα με τις εκτιμήσεις της αγοράς και της φίλιππης κοινότητας .

1.2.1 Συνεισφορά

Η διπλωματική πραγματοποιήθηκε με βάση τα στοιχεία που παρέχει ο ελληνικός ιππόδρομος για τις ιπποδρομίες που διεξήχθησαν στο χώρο του Μαρκόπουλου κατά το έτος 2019. Έπειτα έγινε μια προσπάθεια προσέγγισης των εργαλείων που χρησιμοποιούνται από τους επιστήμονες που εργάζονται για την εξαγωγή προγνωστικών / εκτιμήσεων αλλά και αποδόσεων για τα διάφορα στοιχήματα που αφορούν τα γεγονότα του ιπποδρόμου .

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Συγκέντρωση δεδομένων για τα γεγονότα του ιπποδρόμου κατά το έτος 2019.
2. Μελέτη των μεθόδων πρόβλεψης για αθλητικά γεγονότα.
3. Μετασχηματισμός/Αξιοποίηση των δεδομένων για την προσαρμογή σε μοντέλα πρόβλεψης.
4. Προσαρμογή των δεδομένων σε ένα προτεινόμενο μοντέλο πρόβλεψης αγώνων ιπποδρομίας.
5. Ιδέες για περαιτέρω βελτίωση του μοντέλου πρόβλεψης στο μέλλον.

1.3 Δομή της εργασίας

Η δομή της εργασίας θα κινηθεί τόσο στο βιβλιογραφικό όσο και στο πρακτικό/τεχνικό κομμάτι. Πιο αναλυτικά η οργάνωση μπορεί να ανακεφαλαιωθεί στην εξής μορφή :

Στο Κεφάλαιο 2 θα γίνει μια σύντομη αναφορά στις τεχνικές που έχουν ενσωματωθεί στο οπλοστάσιο της επιστημονικής κοινότητας και προσπαθούν να ρίξουνε φως στο πρόβλημα της πρόβλεψης των αθλητικών γεγονότων αλλά και πιο συγκεκριμένα των ιπποδρομιακών αγώνων . Έπειτα, στο Κεφάλαιο 3 θα αναλυθεί το θεωρητικό υπόβαθρο που διέπει τη μέχρι τώρα προσέγγιση της πρόβλεψη σε στοχαστικά μοντέλα - όπως είναι η έκβαση των αθλητικών γεγονότων και των αγώνων- και τη μεθοδολογία που έχει επικρατήσει σε τέτοιες περιπτώσεις . Στο Κεφάλαιο 4 θα παρουσιαστεί πιο αναλυτικά το πρόβλημα της πρόβλεψης αθλητικών γεγονότων και θα γίνει παράθεση των δεδομένων που έχουν συγκεντρωθεί και αφορούν τις Ελληνικές Ιπποδρομίες κατά το προηγούμενο έτος. Στο κεφάλαιο 5 θα γίνει διαλογή των παραμέτρων που συμβάλουν στη εύρεση καλύτερης λύσης στο πρόβλημα της πρόβλεψης . Στο κεφάλαιο 6 θα παρουσιαστούν οι τεχνικές αξιολόγησης της πρόβλεψης ,ενώ θα φανούν οι δυνατότητες που παρέχει αλλά και τα αδύναμα σημεία του . Ακόμα , στο επόμενο κεφάλαιο θα γίνει παράθεση κάποιων τμημάτων από την υλοποίηση του αλγόριθμου πρόβλεψης και της συμβολής καθεμίας παραμέτρου στο πρόβλημα . Στα κεφάλαια 8 και 9 θα γίνει ανασκόπηση του κύκλου της πρακτικής άσκησης καθώς και των αρμοδιοτήτων που ανέλαβα στην πορεία στο χώρο αυτό. Στο 10ο κεφάλαιο θα γίνει μια συνολική αξιολόγηση και κάποιες προτάσεις για την εταιρεία στην οποία επιτέλεσα την πρακτική άσκηση. Στο 11ο και τελευταίο κεφάλαιο γίνεται μια επισκόπηση της εργασίας και γίνονται προτάσεις που θα αποτελέσουν τροφή για περαιτέρω έρευνα / βελτιστοποίηση του μοντέλου πρόβλεψης που περιγράφεται αναλυτικά και από διαφορετικές οπτικές σε όλη την πορεία της εργασίας.

2 *Το πρόβλημα της πρόβλεψης αθλητικών γεγονότων*

Το επιστημονικό υπόβαθρο γύρω από το οποίο κινείται η πρόβλεψη των νικητών αθλητικών γεγονότων έχει κάνει άλματα, όμως στηρίζεται σε δύο πολύ μεγάλες αλήθειες. Η πρώτη είναι το γεγονός πως η πρόβλεψη επηρεάζεται στη βάση της τυχαιότητας. Αυτό σημαίνει πως ακόμα και οι πιο εμπεριστατωμένες και τεκμηριωμένες προβλέψεις υπάρχει πραγματική πιθανότητα να αποτύχουν να κάνουν αυτό για το οποίο δημιουργήθηκαν-δηλαδή να προβλέψουν το αποτέλεσμα. Επομένως οι κλασσικές μέθοδοι πρόβλεψης που χρησιμοποιούνται σε φαρμακευτικές ή βιομηχανικές επιστήμες θα πρέπει να εμπλουτιστούν και να γίνουν ακόμη πιο πολυκριτηριακές· έτσι ώστε το μαθηματικό μοντέλο να μπορεί να ποσοτικοποιεί την αβεβαιότητα που προκύπτει από κάθε αστάθμητο παράγοντα του αθλητικού γεγονότος.

Τα κλασικά προβλήματα πρόβλεψης κινούνται σε ντετερμινιστική βάση. Αυτό σημαίνει πως όλες οι παράμετροι του προβλήματος λαμβάνονται ως σταθερά μεγέθη· κάτι που μπορεί να μην ισχύει στην πραγματικότητα , όμως κάνει τους υπολογισμούς πολύ πιο εύκολους. Η επιστήμη της στατιστικής εδώ και πάρα πολλά χρόνια έχει δώσει ολοκληρωμένες λύσεις σε προβλήματα πρόβλεψης που είναι πλήρως ντετερμινιστικά. Η ίδια επιτυχία υπάρχει ακόμα και όταν τα προβλήματα είναι στοχαστικά· δηλαδή υπάρχει κάποια αβεβαιότητα για τις μεταβλητές μας. Τα μοντέλα είναι περισσότερο σύνθετα, όμως με παραδοχές για την κατανομή των μεταβλητών μπορούμε και πάλι να κάνουμε αρκετά καλές προβλέψεις. Η πρόβλεψη του αποτελέσματος της ιπποδρομίας θα λέγαμε πως δε θα μπορούσε να εμπίπτει σε κάποια από τις δύο παραπάνω κατηγορίες, καθώς δεν είναι εύκολο να γίνει παραδοχή πως οι μεταβλητές περιγράφονται από κάποια γνωστή μαθηματική κατανομή.

Το δεύτερο γεγονός είναι πως οι συμμετέχοντες δεν έχουν σταθερή απόδοση. Το εκάστοτε μοντέλο πρόβλεψης που θα χρησιμοποιηθεί δε μπορεί να υπολογίσει την ένταση της αντιπαλότητας μεταξύ των ανταγωνιστών και την επίδραση των επιμέρους μεμονωμένων στοιχείων στο αποτέλεσμα των γεγονότων. Οι ανταγωνιστές για τη νίκη περνάνε περιόδους που είναι πιο αποδοτικοί, άλλες που δε βρίσκονται σε τόσο καλή κατάσταση και άλλες που το κίνητρο του αγώνα δίνει μια πολύ διαφορετική εικόνα για την έκβαση του γεγονότος από ότι δίνουν απλά τα στατιστικά στοιχεία πριν την έναρξη του αγώνα.

Η αντιμετώπιση και των δύο παραπάνω ζητημάτων πλέον γίνεται με την εισαγωγή των αλγόριθμων τεχνητής νοημοσύνης (Artificial Intelligence – AI) και της βαθιάς εκμάθησης (Deep Learning).

2.1 Deep Learning

Η βαθιά εκμάθηση (deep learning) αποτελεί έναν σημαντικό κλάδο της μηχανικής μάθησης (machine learning). Είναι ένα πεδίο μελέτης που παρέχει στους υπολογιστές τη δυνατότητα να μαθαίνουν, χωρίς ωστόσο να έχουν προγραμματιστεί ξεκάθαρα για αυτό το σκοπό. Η κεντρική ιδέα είναι πως η επιστήμη της μηχανικής μάθησης κατασκευάζει και μελετά αλγόριθμους που μπορούν να μαθαίνουν από τα δεδομένα και έπειτα να κάνουν προβλέψεις σχετικά με αυτά. Τα μοντέλα δημιουργούνται από πειραματικά δεδομένα και μπορούν να γίνουν προβλέψεις που να βασίζονται στα δεδομένα ή να υποστηριχτεί η λήψη αποφάσεων που εκφράζονται ως το αποτέλεσμα του μοντέλου. Πολλές φορές η μηχανική μάθηση συγχέεται με την υπολογιστική στατιστική και έχει ισχυρούς δεσμούς με τη μαθηματική βελτιστοποίηση, η οποία παρέχει μεθόδους και κλάδους εφαρμογής. Η μηχανική μάθηση εφαρμόζεται σε πολλές υπολογιστικές εργασίες και προτιμάται σε τομείς όπου ο σχεδιασμός και ο κλασσικός προγραμματισμός αλγορίθμων είναι πολύ δύσκολος ή αδύνατος. Οι αλγόριθμοι βαθιάς εκμάθησης περνάνε δεδομένα μέσω διαφόρων "στρωμάτων" αλγορίθμων νευρωνικών δικτύων (neural networks). Κάθε επίπεδο περνά μια πιο απλοποιημένη αναπαράσταση των δεδομένων στο επόμενο στρώμα. Στη βαθιά εκμάθηση, ένα μοντέλο υπολογιστή μαθαίνει να εκτελεί εργασίες ταξινόμησης απευθείας από εικόνες, κείμενο ή ήχο. Τα μοντέλα βαθιάς εκμάθησης μπορούν – όπως και οι άνθρωποι- να μάθουν μέσω αραδειγμάτων να παράγουν όλο και ακριβέστερα αποτελέσματα, ενώ όταν τους παρέχεται ένας πολύ μεγάλος αριθμός δεδομένων μπορούν να επιτύχουν την ακρίβεια της τέχνης, που μερικές φορές υπερβαίνει την αντίστοιχη ανθρώπινη απόδοση.

Η επιστημονική κοινότητα προσπαθεί σύσσωμη να ενσωματώσει τις εφαρμογές της βαθιάς εκμάθησης σε εφαρμογές που μπορούν να εξοικονομήσουν χρόνο και να παράγουν το σωστό αποτέλεσμα χωρίς το ρίσκο της λήψης αποφάσεων από τον ανθρώπινο παράγοντα. Οι πρόσφατες εξελίξεις στη βαθιά εκμάθηση έχουν βελτιωθεί στο σημείο όπου η τελευταία υπερβαίνει τους ανθρώπους σε ορισμένα καθήκοντα όπως η ταξινόμηση αντικειμένων σε εικόνες. Σε συνδυασμό με άλλους τομείς της πληροφορικής που ανεβαίνουν σε δημοτικότητα, όπως η εξόρυξη δεδομένων (data mining) και η τεχνητή νοημοσύνη ανοίγονται νέα πεδία δράσης και έρευνας για την επιστήμη της πληροφορίας και φαίνεται πως το μέλλον τους ανήκει ολοκληρωτικά.

2.2 Εφαρμογή στις προβλέψεις αγώνων

Με τη σωστή χρήση των δεδομένων που υπάρχουν σε ιστορικά γεγονότα και με την τεχνογνωσία στο χώρο της επιστήμης της τεχνολογίας γύρω από την επιστήμη των έξυπνων αλγορίθμων γεννιούνται πολλές δυνατότητες για τους ερευνητές και τους μηχανικούς λογισμικού. Η χρήση των αλγορίθμων βαθιάς εκμάθησης έχει αποδείξει πως θα μπορούσε να δίνει ως αποτέλεσμα καλύτερες προβλέψεις για αγώνες του ιπποδρόμου από ότι και ο πλέον ενημερωμένος άνθρωπος του χώρου.

Από πολλές απόψεις, η αγορά των αγώνων ιπποδρομιών είναι παρόμοια με την χρηματιστηριακή αγορά. Και στις δύο αγορές, οι αποδόσεις από τις επενδύσεις είναι αβέβαιες. Υπάρχουν πολλοί συμμετέχοντες και υπάρχουν πολλές πληροφορίες σχετικά με τις επενδύσεις και τους υπόλοιπους συμμετέχοντες. Αυτό έχει προκαλέσει σημαντικό ενδιαφέρον για τη μελέτη της αποτελεσματικότητας του κέρδους αυτών των αγορών. Στο επόμενο κεφάλαιο θα δούμε τις τεχνικές που έχουν αναπτυχθεί για τον υπολογισμό των πιθανοτήτων επιτυχίας των επενδύσεων και τα συμπεράσματα που μπορούμε να εξάγουμε μέσω της πρόβλεψης της τελικής έκβασης των ιπποδρομιών.

3 *Θεωρητικό υπόβαθρο*

Γενικά, δύο σημαντικές προσεγγίσεις μπορούν να διακριθούν για να προβλέψουν τις πιθανότητες νίκης σε ανταγωνιστικά γεγονότα. Οι εκτιμήσεις των εμπειρογνωμόνων και οι μέθοδοι στατιστικής πρόβλεψης. Έχει αποδειχθεί ότι ακόμα και οι εμπειρογνώμονες έχουν πριν τις εκτιμήσεις τους πληροφορίες που αφορούν την πρόβλεψη του αγώνα. Ωστόσο, είναι γενικά αδύνατο να υπάρχει μακροχρόνια κερδοφορία των προβλέψεων εκτός εάν συνδυαστούν με κάποιο τρόπο με τα μοντέλα πρόβλεψης. Συγκεκριμένα, έχει αποδειχθεί ότι είναι εφικτή η κερδοφορία των συναλλαγών με βάση σύνθετες προβλέψεις από ένα μικρό αριθμό μεμονωμένων αγώνων με διαφορετικά στοιχεία. Τελικά οι απόψεις των εμπειρογνωμόνων μπορούν μόνο να λειτουργήσουν συμπληρωματικά στις προβλέψεις· συνυπολογίζοντας κάποια δεδομένα που δε μπορούν να μπούνε ποσοτικά στο μοντέλο της πρόβλεψης.

Χρησιμοποιήθηκαν αρκετές στατιστικές μέθοδοι για την πρόβλεψη κερδοφόρων πιθανοτήτων σε μεθόδους ανταγωνιστικών γεγονότων, συμπεριλαμβανομένων των μεθόδων τεχνητών νευρωνικών δικτύων με βάση τα δέντρα αποφάσεων και ανάλυση πιθανοτήτων κατά Bayes (Bayesian Inference). Ωστόσο, πολλές από αυτές τις διαδικασίες δεν λαμβάνουν υπόψη την επιρροή του ανταγωνισμού. Για το λόγο αυτό η πιο κοινή προσέγγιση για τη μοντελοποίηση είναι να βλέπουμε τους ανταγωνιστές ως εναλλακτικές λύσεις μέσα σε ένα σύνολο επιλογών και τον νικητή ως συμμετέχοντα του οποίου τα στοιχεία έχουν ως αποτέλεσμα να είναι η προτιμώμενη (δηλαδή, επιλεγμένη) εναλλακτική λύση. Η ερμηνεία αυτή δίνει τη δυνατότητα για εφαρμογή (διακριτών) μοντέλων επιλογής και, πιο ειδικά, τα επικρατέστερα , είναι η λογική παλινδρόμηση η οποία υπολογίζει ρητά τη δύναμη των ανταγωνιστών κατά την εκτίμηση του αποτελέσματος μιας ιπποδρομίας.

3.1 Γενικευμένο γραμμικό μοντέλο

Το μοντέλο Conditional Logit (CL) [McF74] είναι ένα γενικευμένο μοντέλο γραμμικής παλινδρόμησης (regression) μιας λογικής μεταβλητής δύο τιμών και έχει γίνει η συμβατική μέθοδος για την πρόβλεψη των ιπποδρομιών [JJT06] επειδή διατηρεί τις συνδέσεις μεταξύ μεμονωμένων δρομέων σε έναν δεδομένο αγώνα. Δηλαδή, το μοντέλο αυτό μπορεί να εξάγει πληροφορίες από τη σύνθεση ενός αγώνα και συνυπολογίζει τη δυναμικότητα των αντιπάλων κατά την εκτίμηση της πιθανότητας του κάθε ίππου να κερδίσει, ενώ πολλές εναλλακτικές τεχνικές δεν καταφέρνουν να ανταπεξέλθουν στο ανταγωνιστικό στοιχείο εντός του αγώνα. Ωστόσο, παρά αυτό το εννοιολογικό πλεονέκτημα, η CL υφίσταται επίσης σημαντικούς περιορισμούς. Για παράδειγμα, η εκτίμηση βασίζεται σε παραδοχές κατανομών, οι οποίες μπορεί να παραβιαστούν στην πρακτική εφαρμογή. Επιπλέον, δεν είναι σε θέση να μοντελοποιήσει μη γραμμικές σχέσεις μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών, εκτός εάν αυτές καθορίζονται ρητά από τον δημιουργό του μοντέλου.

Για την αντιμετώπιση αυτών των προβλημάτων χρειάστηκε η συμβολή του Support Vector Machine (SVM) (Ede07, LSJ08) Το SVM δε βασίζεται σε υποθέσεις κατανομής και μπορεί να προσεγγίσει κάθε μη γραμμική συνάρτηση με αρκετή ακρίβεια. Ωστόσο, τα Support Vector Machines είναι διαδικασίες μαύρου κουτιού (black box) και δεν επιτρέπουν προβλέψεις ως προς τη συμβολή των μεμονωμένων μεταβλητών που πρέπει να διακρίνονται. Επιπλέον, λόγω της υπολογιστικής πολυπλοκότητας που σχετίζονται με την κατασκευή προτύπων πρόβλεψης βασισμένων σε SVM, είναι

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

δύσκολο να προσαρμοστεί το μοντέλο σε μεγάλα σύνολα δεδομένων που συναντώνται συχνά στο χώρο των αγώνων των ιπποδρομιών.

Η παλινδρόμηση CL έχει εφαρμοστεί με επιτυχία σε πολλές εφαρμογές μάρκετινγκ για να μοντελοποιηθεί [CS82] και εισήχθη ως κατάλληλη τεχνική για την πρόβλεψη αποτελεσμάτων ιπποδρομιών από τους Bolton & Chapman [BC86]. Διακριτικό του πλεονέκτημα για αυτό το καθήκον πηγάζει από το γεγονός ότι η CL διατηρεί τις συνδέσεις εντός της εναλλακτικές λύσεις ενός συνόλου επιλογών (δηλαδή μεταξύ ανταγωνιστών σε έναν αγώνα). Δηλαδή, η CL επιτρέπει τη νίκη η πιθανότητα ενός αλόγου να εκτιμηθεί σε συνάρτηση με τις πιθανότητες ενός άλλου αλόγου και, συνεπώς, αντιπροσωπεύει τον ανταγωνισμό εντός του αγώνα.

Έστω

$$S = \left\{ \boldsymbol{x}_i^j, y_i^j \right\}_{i=1, j=1}^{n_j, R}$$

συμβολίζει ένα σετ δεδομένων (dataset) από R προηγούμενες κούρσες , όπου $x_i^j \in R^m$ αντιπροσωπεύει ένα άλογο i σε μια κούρσα j που χαρακτηρίζεται από m ανεξάρτητες μεταβλητές και n_j είναι ο αριθμός των αλόγων στην κούρσα j . Ο στόχος του μοντέλου πρόβλεψης CL είναι να υπολογίσει ένα διάνυσμα πιθανοτήτων νίκης $p^j = (p_1^j , p_2^j, \dots p_n^j)$ για τον αγώνα j . Τα p_i^j συμβολίζουν την εκτιμώμενη πιθανότητα που δίνει το μοντέλο το άλογο i να κερδίσει τον αγώνα j. Έστω W_i^j μια δυαδική μεταβλητή - δείκτης (νίκη/ήττα) που προσδιορίζεται ως

$$W_i^j = \begin{cases} 1 \text{ if } \boldsymbol{\beta} \cdot \boldsymbol{x}_i^j + \varepsilon_i^j > \boldsymbol{\beta} \cdot \boldsymbol{x}_l^j + \varepsilon_l^j \quad \forall i, l = 1, .., n_j, i \neq l \\ 0 \text{ otherwise,} \end{cases}$$

Όπου β είναι το διάνυσμα των σταθερών συντελεστών που μετράνε τη σχετική κατανομή των ανεξάρτητων μεταβλητών που περιέχονται στο διάνυσμα x_i^j και περιγράφουν τον ίππο i στον αγώνα j ενώ ο όρος σφάλματος ε_i^j συμπεριλαμβάνει τη πληροφορία που δεν έχει συμπεριληφθεί. Με μερικές ακόμα παραδοχές μπορούμε να καταλήξουμε πως η πιθανότητα του αλόγου i να κερδίσει την κούρσα j δίνεται από την ακόλουθη συνάρτηση :

$$p_i^j = \frac{\exp\left(\boldsymbol{\beta} \cdot \boldsymbol{x}_i^j\right)}{\sum_{i=1}^{n_j} \exp\left(\boldsymbol{\beta} \cdot \boldsymbol{x}_i^j\right)}.$$

Το διάνυσμα των σταθερών του προβλήματος , **β**, υπολογίζεται με τη μέθοδο της εκτίμησης της μέγιστης πιθανότητας (Maximum Likelihood Estimation – MLE)

$$\boldsymbol{\beta} \leftarrow \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \prod_{j=1}^{R} p_{i*}^j = \prod_{j=1}^{R} \exp\left(\boldsymbol{\beta} \cdot \boldsymbol{x}_{i*}^j\right) \Big/ \sum_{i=1}^{m_j} \exp\left(\boldsymbol{\beta} \cdot \boldsymbol{x}_i^j\right),$$

Όπου x_i^j * αντιπροσωπεύει τον νικητή της κούρσας j

Εδώ είναι σημαντικό να σημειωθεί πως οι σταθερές του μοντέλου, β, υπολογίζονται με τη μεγιστοποίηση του συντελεστή πιθανότητας σε ολόκληρο το σετ δεδομένων, χωρίς να εξετάζεται ποιοι συμμετέχοντες – ίπποι - ανταγωνίζονται σε ένα δεδομένο αγώνα. Η περαιτέρω ανάλυση ωστόσο , της τελικής μορφής του παραπάνω διανύσματος ξεφεύγει από τα πλαίσια της παρούσας εργασίας .

Με άλλα λόγια, υπολογίζεται η ατομική πιθανότητα ενός αλόγου να είναι νικητής (σε οποιαδήποτε κούρσα). Επομένως, μπορεί κανείς να υποστηρίξει ότι αυτό το μοντέλο προσέγγισης υποθέτει ότι τα νικητήρια άλογα είναι εννοιολογικά διαφορετικά από αυτά που δεν κερδίζουν και ότι όλες οι (διαθέσιμες) πληροφορίες που θα μπορούσαν ενδεχομένως να επηρεάσουν τις πιθανότητες

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

όλων των συμμετεχόντων περιέχονται στις ανεξάρτητες μεταβλητές. Η παραδοχή αυτή δεν ισχύει σε ανταγωνιστικά σενάρια -όπως είναι οι ιπποδρομίες . Η πιθανότητα ενός ίππου να κερδίσει εξαρτάται σαφώς και από την κατάσταση των ανταγωνιστών του. Ως εκ τούτου, είναι προτιμότερο να προβλεφθεί η πιθανότητα νίκης σε σχέση με τους αντιπάλους – όπως γίνεται στο μοντέλο CL.

Παρά τα αρκετά πλεονέκτημα της CL όσον αφορά τις προβλέψεις υποφέρει από μερικά σημαντικά μειονεκτήματα. Ειδικότερα, η εφαρμογή της CL προϋποθέτει ότι υπάρχουν σφάλματα κατανεμημένα σύμφωνα με τη διπλή εκθετική κατανομή, και αυτό είναι απίθανο στην πράξη [GHB03]. Επιπλέον, το μοντέλο CL δεν είναι σε θέση να υπολογίσει μη γραμμικές αλληλεπιδράσεις μεταξύ ανεξάρτητων μεταβλητών και μεταξύ ανεξάρτητων μεταβλητών και του στόχου. Δηλαδή, η σταθμισμένη· που εκφράζεται μέσω του **β**· επίδραση της ανεξάρτητης μεταβλητής στο στόχο είναι γραμμική και δρα αθροιστικά. Επιπλέον, τα επιχειρήματα από τη θεωρία της στατιστικής μάθησης [Vap95] δείχνουν ότι η εκτίμηση των συντελεστών βάσει της μέγιστης πιθανότητας (MLE) μπορεί να είναι ασταθής εάν υπάρχει μεγάλος αριθμός από ανεξάρτητες μεταβλητές, όπως συμβαίνει συνήθως στους αγώνες ιπποδρομίας.

3.2 Μέθοδος του τυχαίου δάσους

Η διαδικασία του τυχαίου δάσους (Random Forest - RF) αποτελεί μια υπερσύγχρονη μέθοδο εκμάθησης μηχανών που προτάθηκε για πρώτη φορά το 2001. Έχει χαρακτηριστικά που επιτρέπουν να ξεπεραστούν οι περιορισμοί τόσο των CL όσο και των SVM μοντέλων πρόβλεψης: Πρώτον, η ικανότητά του να σχηματίζει πολύ ακριβείς προβλέψεις έχει επιβεβαιώθηκε σε διάφορες εμπειρικές μελέτες (π.χ. CN06 , CLB04 ,HL05,LSJ08). Δεύτερον οι μέθοδοι RF μπορούν να αναγνωρίσουν τη σημασία των μεμονωμένων μεταβλητών και είναι ικανές να διακρίνουν μη γραμμικές αλληλεπιδράσεις μεταξύ μεταβλητών. Τρίτον, η διαδικασία RF είναι γρήγορη και εφαρμόζεται σε μεγάλη κλίμακα, σε σημείο που μεγάλα σύνολα δεδομένων μπορούν να αντιμετωπιστούν με ευκολία. Τέλος, θα μπορούσε επίσης να υποτεθεί ότι η μέθοδος του τυχαίου δάσους είναι κατάλληλη για την πρόβλεψη των αποτελεσμάτων των γεγονότων που σχετίζονται με τις αγορές στοιχημάτων (π.χ., ιπποδρομίες), καθώς υπάρχουν ομοιότητες μεταξύ του τρόπου με τον οποίο οι αγορές RF και οι αγορές στοιχημάτων συγκεντρωτικές πληροφορίες.

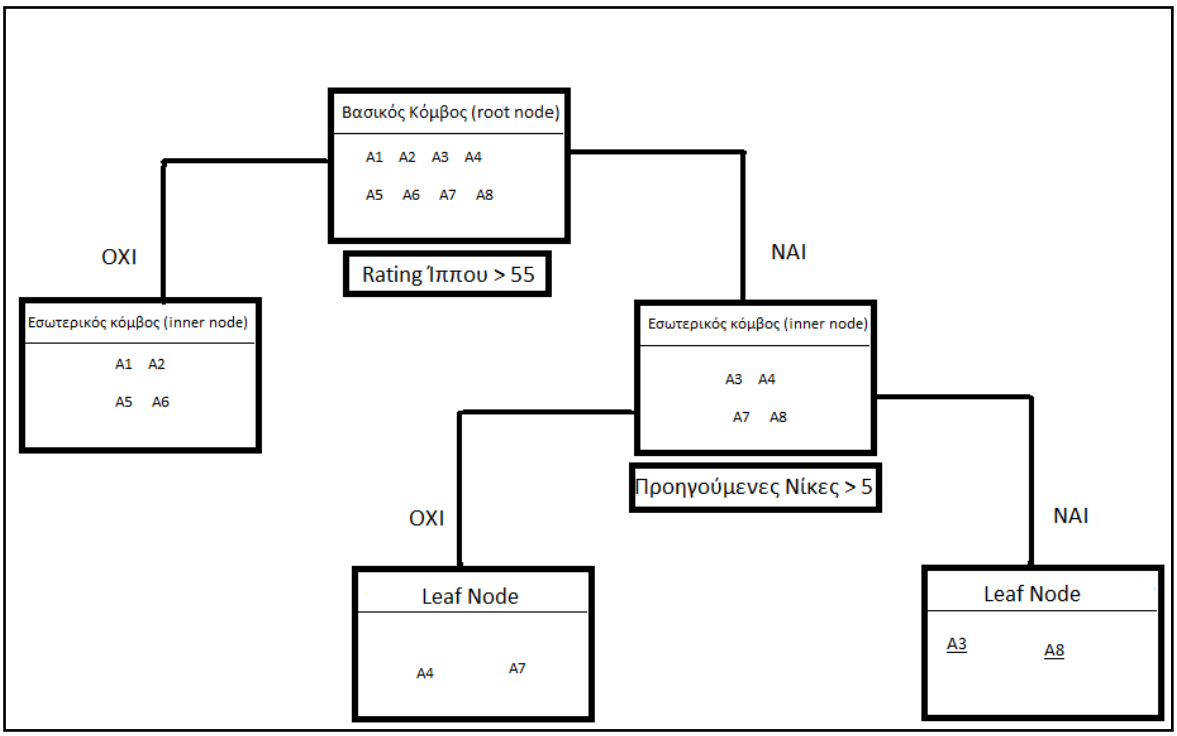
Με άλλα λόγια, η RF είναι μια μη παραμετρική τεχνική που δεν βασίζεται σε καμία υπόθεση πως τα δεδομένα ακολουθούν οποιαδήποτε κατανομή και είναι σε θέση να εξάγει αυτόματα μη γραμμικές σχέσεις [Bre01]. Επιπλέον, η μέθοδος RF είναι σε θέση να παρέχει κατανοητές εκτιμήσεις της σημασίας κάθε μεταβλητής, ενώ οι περισσότερες εναλλακτικές μη γραμμικές μέθοδοι, όπως τα τεχνητά νευρωνικά δίκτυα ή το SVM λειτουργούν ως μαύρα κουτιά. Η RF είναι επίσης προτιμότερη από τις εναλλακτικές από υπολογιστικής άποψης. Συγκεκριμένα, η δομή του μοντέλου εξασφαλίζει ταχύτητα και μπορεί να οργανωθεί εύκολα με παράλληλο τρόπο.

Από την άλλη πλευρά, τα πρότυπα RF περιορίζονται σε άλλους τομείς. Δεν είναι σε θέση να εισάγουν τον ανταγωνισμό που υπάρχει κατά τη διάρκεια του γεγονότος. Συνεπώς, η RF μπορεί τελικά να θεωρηθεί ως προσέγγιση συμπλήρωσης και όχι αντικατάστασης του γραμμικού μοντέλου CL των ιπποδρομιών. Επομένως, ο γενικός κανόνας είναι πως οι δύο μέθοδοι συνδυάζονται έτσι ώστε να πάρουμε όσο το δυνατόν καλύτερα αποτελέσματα στην τελική πρόβλεψη.

Το τυχαίο δάσος βασίζεται στη γνωστή μεθοδολογία CART {Classification and Regression Trees} (ταξινόμηση και δένδρα παλινδρόμησης) [BFO84] για την κατασκευή μοντέλων δέντρων αποφάσεων. Τέτοια μοντέλα περιλαμβάνουν αναδρομικό διαχωρισμό του συνόλου δεδομένων S σε ξεχωριστά παραδείγματα (δηλαδή άλογα) διαφορετικών ομάδων. Αυτή η αρχή απεικονίζεται στο παρακάτω σχήμα χρησιμοποιώντας τεχνητά δεδομένα δύο αγώνων τεσσάρων δρομέων, όπου τα άλογα χαρακτηρίζονται από δύο ανεξάρτητες μεταβλητές , όπως φαίνεται στον πίνακα παρακάτω

ΚΟΥΡΣΑ 1				
ΟΝΟΜΑ	**ΚΩΔΙΚΟΣ ΙΠΠΟΥ**	**ΠΡΟΗΓΟΥΜΕΝΕΣ ΝΙΚΕΣ**	**RATING***	**ΝΙΚΗΤΗΣ****
ΑΣΠΡΟΥΛΗΣ	A1	2	51	0
ΤΕΪΚ ΜΙ ΤΟΥ	A2	3	44	0
ΜΥΚΟΝΙΑΤΙΣΣΑ	A3	6	57	1
ΣΤΕΪ ΧΕΒΙ	A4	4	50	0
ΚΟΥΡΣΑ 2				
ΟΝΟΜΑ	**ΚΩΔΙΚΟΣ ΙΠΠΟΥ**	**ΠΡΟΗΓΟΥΜΕΝΕΣ ΝΙΚΕΣ**	**RATING***	**ΝΙΚΗΤΗΣ****
ΛΙΤΛ ΠΡΙΝΣΕΣ ΜΠΑΪ ΔΕ	A5	1	33	0
ΜΠΟΥΚ	A6	4	51	0
ΤΙΜΠΟΥΚΤΟΥ	A7	7	54	0
ΜΠΙΓΚ ΜΙΡΑΚΛ	A8	5	59	1

* Το Rating του Ίππου είναι μια τιμή που του δίνεται βάσει κάποιων παραμέτρων που συνυπολογίζονται από αποτελέσματα προηγούμενων αγώνων . Γίνεται αναλυτικά αναφορά σε παρακάτω κεφάλαιο.
** Νικητής είναι μια δυαδική μεταβλητή (binary variable). Παίρνει τιμή 1 για το νικητή ίππο και τιμή 0 για τους ίππους που δεν είναι
νικητές αν

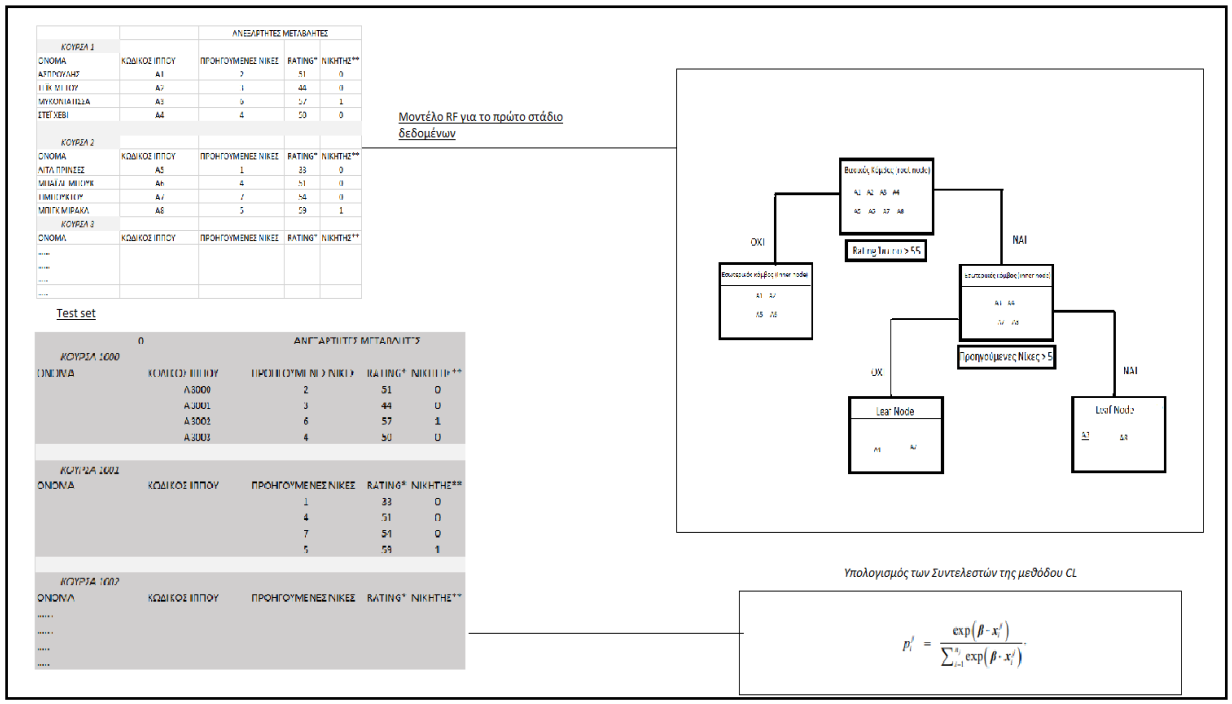


Απεικόνιση της προσέγγισης των δέντρων αποφάσεων για την κατασκευή προτύπων πρόγνωσης για διακριτική ανάλυση χρησιμοποιώντας τεχνητά δεδομένα από δύο αγώνες με τέσσερις συμμετέχοντες στον καθένα . Κάθε δρομέας χαρακτηρίζεται από δύο ανεξάρτητες μεταβλητές. Η δυαδική εξαρτώμενη μεταβλητή δηλώνεται από τον νικητή και η υπογραμμισμένη γραμματοσειρά χρησιμοποιείται για να τονίσει το άλογο που τελειώνει πρώτο σε κάθε αγώνα.

Το μοντέλο τυχαίου δάσους διαμόρφωσης δύο σταδίων (2 Stage Random Forest -2RF) προσφέρει δύο σημαντικά πλεονεκτήματα σε σχέση με ένα μοντέλο πρόβλεψης σε ένα στάδιο: Πρώτον, η ανάγκη για τον υπολογισμό του ανταγωνισμού εντός του αγώνα αναβάλλεται για το δεύτερο στάδιο και μπορούν να χρησιμοποιηθούν αρκετοί αλγόριθμοι πρόβλεψης για τη σύλληψη των εξελιγμένων σχέσεων μεταξύ θεμελιωδών μεταβλητών που αφορούν την ικανότητα του αλόγου και το αποτέλεσμα των αγώνων στο πρώτο στάδιο. Για παράδειγμα, ο Edelman [Ede07]πρότεινε ένα μοντέλο παλινδρόμησης των διανυσμάτων υποστήριξης (SVM) για να επεξεργαστεί έναν μεγάλο αριθμό υψηλά συσχετισμένων μεταβλητών (όπως γενικά συναντώνται στις προβλέψεις ιπποδρομιών) έτσι ώστε να αποκαλύψει μη γραμμικές σχέσεις μεταξύ των μεταβλητών εισόδου. Παρομοίως, ο Lessmann [LSJ08] κατέδειξε ότι τα μοντέλα ταξινόμησης SVM, τα οποία κατηγοριοποιούν τα άλογα σε νικητές και ηττημένους, αποτελούν μια ελπιδοφόρα εναλλακτική λύση στην παλινδρόμηση. Δεύτερον, ο διαχωρισμός των θεμελιωδών μεταβλητών και των στοιχηματικών αποδόσεων της αγοράς έχει αποδειχθεί ότι είναι αποτελεσματική [SJ07] . Ένα μοντέλο που συνδυάζει απλώς τις θεμελιώδεις μεταβλητές και τις αποδόσεις της αγοράς σε πρώτο στάδιο θα ήταν προσανατολισμένο κυρίως στις αποδόσεις της αγοράς και όλες οι άλλες ανεξάρτητες μεταβλητές θα λάμβαναν μικρότερη προσοχή. Αυτό είναι ανεπιθύμητο, διότι οι αποδόσεις της αγοράς αντανακλούν τη γνώμη του κοινού σχετικά με τη νίκη του αλόγου και σε σημαντικό βαθμό περιλαμβάνουν τις διαθέσιμες στο κοινό πληροφορίες, ενώ ένα μεγάλο μέρος των πληροφοριών μπορεί να μην είναι διαθέσιμες στο κοινό που στοιχηματίζει και μεταβάλλει με αυτό τον τρόπο τις αποδόσεις.

Ένα ικανοποιητικό μοντέλο συνδυάζει τις μεθόδους που περιγράφονται παραπάνω (CL και RF), αλλά χρησιμοποιεί τη μεθοδολογία RF στο πρώτο στάδιο για τους λόγους που περιγράφονται παραπάνω. Επιπλέον, υποτίθεται ότι η ενσωμάτωση διαφόρων μοντέλων βάσης, όπως δηλώνεται από το πλαίσιο RF, καθιστά τα αντίστοιχα μοντέλα ιδιαίτερα κατάλληλα για την πρόβλεψη των ιπποδρομιών. Αυτό οφείλεται στο γεγονός ότι το πρώτο στάδιο μιμείται, σε κάποιο βαθμό, τις διαδικασίες λήψης αποφάσεων που τελικά οδηγούν στον σχηματισμό αποδόσεων στην αγορά. Δηλαδή, ένα δεδομένο άλογο αξιολογείται από διάφορους παίκτες με διαφορετικά επίπεδα εμπειρίας (παρόμοια με μεμονωμένα μοντέλα CART σε RF). Οι παίκτες, κατά πάσα πιθανότητα, θεωρούν διαφορετικά σύνολα διαθέσιμων στο κοινό πληροφοριών (διαφορετικές ανεξάρτητες μεταβλητές) κατά τη λήψη των αποφάσεών τους. Στη συνέχεια τοποθετούν τα στοιχήματά τους (ψηφοφορία) με βάση την εκτίμηση της ικανότητας του αλόγου. Η συγκέντρωση αυτών των εκτιμήσεων παράγει τις τιμές της αγοράς.

Σε δεύτερο στάδιο , εφόσον έχει γίνει μια πρώτη ταξινόμηση από το RF οι συμμετέχοντες που έχουν αναδειχθεί οι πιθανοί νικητές μπορούν να ¨ανταγωνιστούν¨ μεταξύ τους με το μοντέλο πρόβλεψης CL, το οποίο αναπτύχθηκε παραπάνω.



Αρχιτεκτονική του μοντέλου πρόβλεψης RF σε δύο στάδια για την πρόβλεψη του αποτελέσματος των ανταγωνιστικών γεγονότων, υποθέτοντας ένα σύνολο δεδομένων που περιέχει 1000 αγώνες και χρησιμοποιώντας τον ίδιο αριθμό αγώνων για το πρώτο και το δεύτερο στάδιο μοντελοποίησης αντίστοιχα.

4 *Προετοιμασία και ανάλυση των ιστορικών δεδομένων ιπποδρομιακών αγώνων*

Η παρούσα εργασία εστιάζει στη δημιουργία ενός μοντέλου πρόβλεψης , στο πλαίσιο που παρουσιάστηκε αναλυτικά παραπάνω σε θεωρητικό υπόβαθρο. Για να φτάσουμε όμως σε αυτό το σημείο θα πρέπει να δώσουμε βάση σε κάποια ζητήματα που στέκονται εμπόδιο πριν την σύλληψη και την ανάλυση του μοντέλου πρόβλεψης

- ***Προσδιορισμός των κρίσιμων ανεξάρτητων μεταβλητών που τεκμηριωμένα έχουν επίδραση στο αποτέλεσμα του αγώνα***. Κανένα μοντέλο δε μπορεί να λάβει υπόψιν όλους τους παράγοντες που μπορούν να συνεισφέρουν στην τελική κατάταξη του αγώνα, αλλά υπάρχουν και κάποιες παράμετροι που ενώ είναι διαθέσιμοι σε στατιστικά στοιχεία δε ξέρουμε κατά πόσο αυτοί επηρεάζουν το αποτέλεσμα. Σε πολλές περιπτώσεις τα διαθέσιμα στατιστικά δε ξέρουμε εάν είναι όντως χρήσιμα ή απλά θόρυβος για το μοντέλο μας. Για παράδειγμα το χρώμα του ίππου είναι ένα διαθέσιμο στατιστικό και κάτι που υπάρχει ως στατιστικό στοιχείο σε μια ιπποδρομία. Επηρεάζει το αποτέλεσμα της ιπποδρομίας; Μάλλον όχι. Ο αναβάτης του ίππου επηρεάζει το αποτέλεσμα της ιπποδρομίας; Μάλλον ναι.

  Σε κάθε περίπτωση θα πρέπει με κάποιον τρόπο να βρούμε ποιες μεταβλητές είναι αυτές που έχουν συσχέτιση τόσο μεταξύ τους όσο και με το αποτέλεσμα του αγώνα.

- ***Συγκέντρωση και ένωση των διαθέσιμων δεδομένων.*** Το πιο κρίσιμο σημείο για να ξεκινήσουμε το μοντέλο μας είναι να συγκεντρώσουμε τα δεδομένα από περασμένες ιπποδρομίες και με τα εργαλεία που έχουμε στη διάθεση μας να τα ενοποιήσουμε σε ένα φύλλο δεδομένων. Έπειτα, θα μπορέσουμε με τη χρήση του κατάλληλου εργαλείου να πάρουμε αποφάσεις για το τι θα χρειαστεί στον αλγόριθμό μας και τι δεν είναι απαραίτητο για την ανάλυσή μας.

4.1 Διαθέσιμα ιστορικά στοιχεία

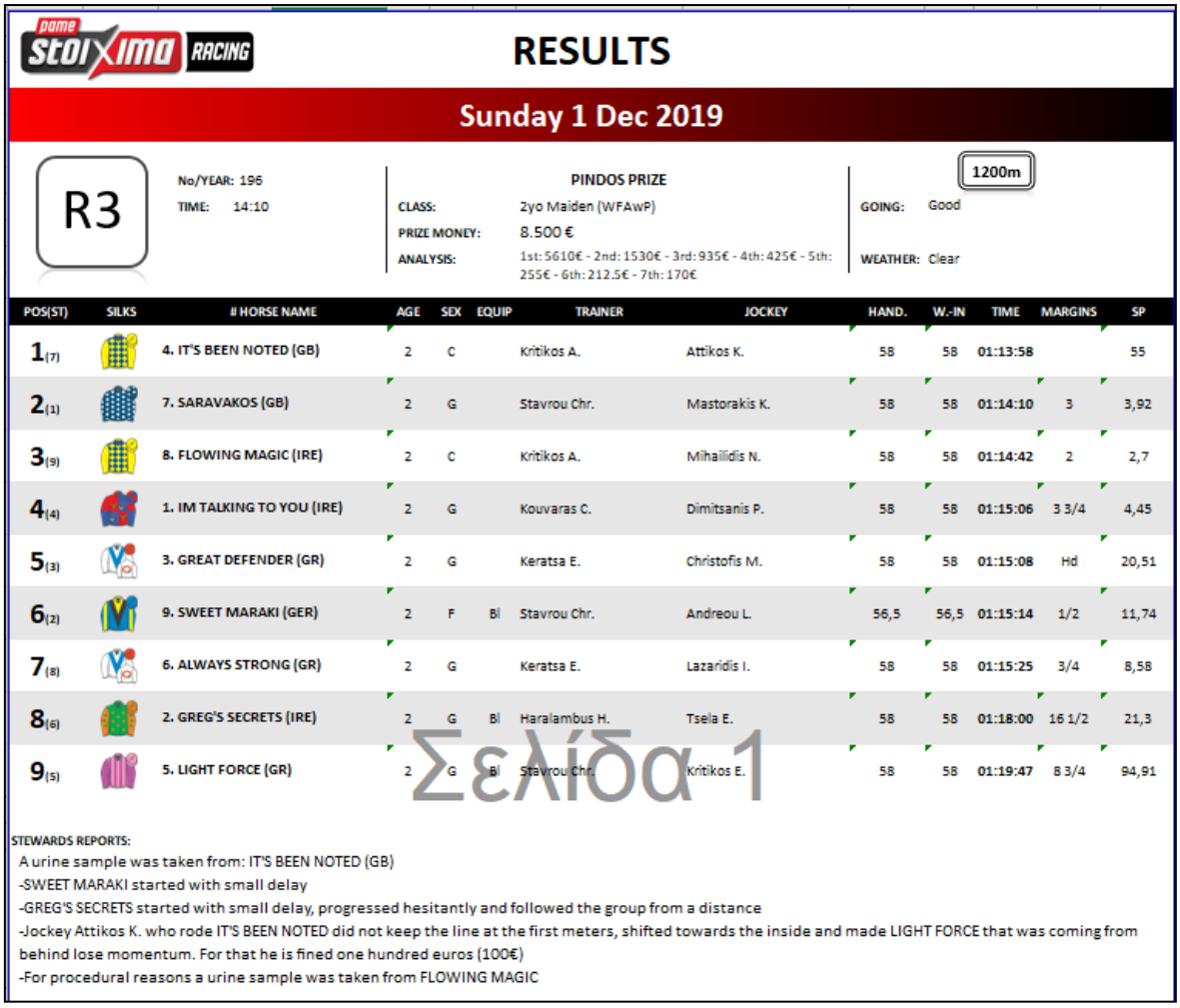
Τα ιστορικά στοιχεία που υπάρχουν θα πρέπει να έρθουν σε μια μορφή που θα μπορούν να αξιοποιηθούν από τα εργαλεία που θα επιλέξουμε για την ανάλυση των δεδομένων μας. Για αρχή έχουμε διαθέσιμα τα αποτελέσματα των ιπποδρομιών που διεξήχθησαν στο Μαρκόπουλο κατά το έτος 2019. Συνολικά σε ένα χρόνο αγωνίστηκαν 310 ίπποι ηλικίας από 3 έως 14 ετών σε 235 ιπποδρομίες. Τα στοιχεία μας διαμορφώνονται ως εξής:

- Ημερομηνία αγώνων
- Καιρός κατά τη διάρκεια της κούρσας
- Μήκος διαδρομής (σε μέτρα)
- Έπαθλο αγώνα
- Πλήθος συμμετεχόντων ίππων
- Ονόματα συμμετεχόντων ίππων
- Φύλο ίππου
- Ηλικία ίππου
- Βάρος ίππου
- Επιπλέον εξοπλισμός (πχ παρωπίδες)
- Ονόματα αναβατών κάθε ίππου
- Ονόματα προπονητών κάθε ίππου
- Απόσταση τερματισμού από τον επόμενο ίππο
- Rating ίππου



Στιγμιότυπο από το «φώτο φίνις» αγώνα ιπποδρομίας την Κυριακή 19/05/2019 στο Μαρκόπουλο Αττικής.

Ένα σημαντικό πλήθος από τα παραπάνω στοιχεία μπορεί να βρεθεί στο αρχείο των Ελληνικών Ιπποδρομιών για κάθε χρήστη που έχει πρόσβαση το ίντερνετ. Αναλυτικά τα αποτελέσματα των ιπποδρομιών είναι στην ιστοσελίδα horseraces.pamestoixima.gr



RESULTS

Sunday 1 Dec 2019

R3

No/YEAR: 196
TIME: 14:10

PINDOS PRIZE
CLASS: 2yo Maiden (WFAwP)
PRIZE MONEY: 8.500 €
ANALYSIS: 1st: 5610€ - 2nd: 1530€ - 3rd: 935€ - 4th: 425€ - 5th: 255€ - 6th: 212.5€ - 7th: 170€

1200m
GOING: Good
WEATHER: Clear

POS(ST)	SILKS	# HORSE NAME	AGE	SEX	EQUIP	TRAINER	JOCKEY	HAND.	W.-IN	TIME	MARGINS	SP
1 (7)		4. IT'S BEEN NOTED (GB)	2	C		Kritikos A.	Attikos K.	58	58	01:13:58		55
2 (1)		7. SARAVAKOS (GB)	2	G		Stavrou Chr.	Mastorakis K.	58	58	01:14:10	3	3,92
3 (9)		8. FLOWING MAGIC (IRE)	2	C		Kritikos A.	Mihaïlidis N.	58	58	01:14:42	2	2,7
4 (4)		1. IM TALKING TO YOU (IRE)	2	G		Kouvaras C.	Dimitsanis P.	58	58	01:15:06	3 3/4	4,45
5 (3)		3. GREAT DEFENDER (GR)	2	G		Keratsa E.	Christofis M.	58	58	01:15:08	Hd	20,51
6 (2)		9. SWEET MARAKI (GER)	2	F	Bl	Stavrou Chr.	Andreou L.	56,5	56,5	01:15:14	1/2	11,74
7 (8)		6. ALWAYS STRONG (GR)	2	G		Keratsa E.	Lazaridis I.	58	58	01:15:25	3/4	8,58
8 (6)		2. GREG'S SECRETS (IRE)	2	G	Bl	Haralambus H.	Tsela E.	58	58	01:18:00	16 1/2	21,3
9 (5)		5. LIGHT FORCE (GR)	2	G	Bl	Stavrou Chr.	Kritikos E.	58	58	01:19:47	8 3/4	94,91

Σελίδα 1

STEWARDS REPORTS:
A urine sample was taken from: IT'S BEEN NOTED (GB)
-SWEET MARAKI started with small delay
-GREG'S SECRETS started with small delay, progressed hesitantly and followed the group from a distance
-Jockey Attikos K. who rode IT'S BEEN NOTED did not keep the line at the first meters, shifted towards the inside and made LIGHT FORCE that was coming from behind lose momentum. For that he is fined one hundred euros (100€)
-For procedural reasons a urine sample was taken from FLOWING MAGIC

Στιγμιότυπο από τα αποτελέσματα της ιπποδρομίας που διεξήχθη την Κυριακή 01/12/2019 στο Μαρκόπουλο Αττικής

4.2 Διαλογή των δεδομένων

Προτού γίνει επιλογή των μεταβλητών που είναι κρίσιμες για το πρόβλημα θα πρέπει γίνει μια ανασκόπηση των δεδομένων που έχουμε συλλέξει. Ως αφετηρία θα πάρουμε κάποια αρχεία xlsx (MS Excel) τα οποία έχουμε φτιάξει και περιέχουν.

- Ένα αρχείο με τα ονόματα των Ίππων, την καταγωγή τους (χώρα), την ηλικία τους και το φύλο τους.
- Ένα αρχείο με τις κλάσεις (κατηγορίες) στο οποίο χωρίζονται οι ίπποι για να διαγωνιστούν
- Ένα αρχείο με τις σημειώσεις των αγωνοδικών που καταγράφουν γεγονότα τα οποία πέφτουν στην αντίληψή τους και ελέγχους ντόπινγκ που πραγματοποιούνται.
- Ένα αρχείο με τα αποτελέσματα της κάθε ιπποδρομίας μέσα στο 2019 (και αναλυτικά η κατάταξη των ίππων)
- Ένα αρχείο με τις αποχές των αλόγων (ασθένειες τραυματισμοί , διάστημα αποχής)
- Ένα αρχείο με τα ratings των αλόγων.

Το rating (αξιολόγηση) των αλόγων είναι το σύνολο της βαθμολογίας που παίρνουμε βάση την επίδοση τους στους αγώνες. Δεν υπάρχει κάποιος τύπος που να υπολογίζει το rating, αλλά αυτό γίνεται συνήθως στα πλαίσια των ιπποδρομιών από τον λεγόμενο ισοζυγιστή.

Η δουλειά του ισοζυγιστή είναι να αξιολογεί την απόδοση των ίππων που έτρεξαν στην ιπποδρομία διαλέγοντας έναν οδηγό ίππο ως τον έναν που θα είναι το σημείο αναφοράς και για τους υπόλοιπους διαγωνιζόμενους. Γενικά επηρεάζεται από αντικειμενικά κριτήρια (πχ διαφορά νίκης – βάρος ίππου) αλλά συνυπολογίζει και σημάδια που υπάρχουν στην κούρσα και δεν γίνονται αντιληπτά από τα μοντέλα πρόβλεψης. Για παράδειγμα ένας αναβάτης που αφήνει στα τελευταία μέτρα τον ίππο να καλπάσει χωρίς πίεση θα χάσει πιθανόν τη μεγάλη διαφορά από τους υπόλοιπους. Σε αυτή την περίπτωση ο ισοζυγιστής που παρακολουθεί την κούρσα μπορεί να μη διαλέξει να συνυπολογίσει τη διαφορά νίκης για την επανεκτίμηση του rating του αλόγου.

Αν και δεν αποτελεί αντικειμενικό κριτήριο, το rating είναι ένα από τα βασικά μεγέθη που συγκαταλέγονται στην πρόβλεψη των ιπποδρομιών και δίνει ένα γενικό πλαίσιο βάσει του οποίου οι ίπποι κατατάσσονται σε διάφορες κατηγορίες. Θα λέγαμε πως το rating αποτελεί το βασικό μετρήσιμο στοιχείο για την απόδοση του κάθε αλόγου και για αυτό σε πρώτο στάδιο θα πρέπει να το χρησιμοποιήσουμε στους υπολογισμούς μας ως βασική ανεξάρτητη μεταβλητή. Τα φύλλα που έχουμε στη διάθεση μας περιέχουν τα ratings από όλα τα άλογα και τις αλλαγές που έχουν γίνει στην αξιολόγηση κατά τη διάρκεια του έτους. Σημειώνουμε σε αυτό το σημείο πως το rating μετά από κάθε αγώνα μπορεί να αλλάξει αλλά μπορεί να παραμείνει και το ίδιο. Πάντοτε γίνεται αυτό που κρίνει πως πρέπει ο ισοζυγιστής.

2Η ΙΠΠΟΔΡΟΜΙΑ / #98

ΑΦ(ΣΤ)	ΧΡΩΜΑΤΑ	# ΟΝΟΜΑ ΙΠΠΟΥ	ΗΛΙΚΙΑ	ΓΕΝΟΣ	ΕΞΟΠ.	ΠΡΟΠΟΝΗΤΗΣ	ΑΝΑΒΑΤΗΣ	ΧΑΝΤ.	ΜΕΤΑΖ.	ΧΡΟΝΟΣ	ΜΗΚΗ	ΑΠΟΔΟΣΗ
1(5)		3. ΜΙΣ ΠΕΓΚΥ (GR)	5	Φορ.	Π	Τζαβέλας Π.	Τσέλα Ε.	59,5	59,5	**00:59:52**		12,19
2(4)		8. ΜΙΣ ΡΟΣΙΝΑ (IRE)	5	Φορ.	Π	Καμούτσης Γ.	Λαζαρίδης Ι.	54	54	**00:59:60**	1/2	3,96
3(2)		2. ΓΚΡΕΪΣ ΔΕ ΦΛΟΡ (IRE)	4	Φορ.	Π	Χαραλάμπους Χ.	Δρόσος Δ.	59,5	59,5	**00:59:76**	1	5,74
4(6)		5. ΣΑΦΦΑΧ (USA)	4	Φορ.		Σταύρου Χρ.	Δημητσάνης Π.	58	58	**01:00:14**	2 1/4	5,04
5(1)		6. ΠΙΚΑΝΤΙΛΟ (GB)	7	Φορ.		Μαυρόπουλος Σπ.	Χριστοφής Μ.	56	56	**01:00:37**	1 1/4	18,49
6(7)		7. ΜΟΓΓΥ (USA)	4	Φορ.	Π	Χαραλάμπους Χ.	Καποδίστριας Κ.	55	55	**01:00:75**	2 1/4	67,68
7(3)		1. ΓΚΟΣΤ (GB)	4	Φορ.	Π	Κρητικός Α.	Μιχαηλίδης Ν.	60	60	**01:01:05**	2 3/4	9,59
		4. ΜΠΕΪΜΠΥ ΣΑΟΥΕΡ (GB)					ΔΙΑΓΡΑΦΗ					

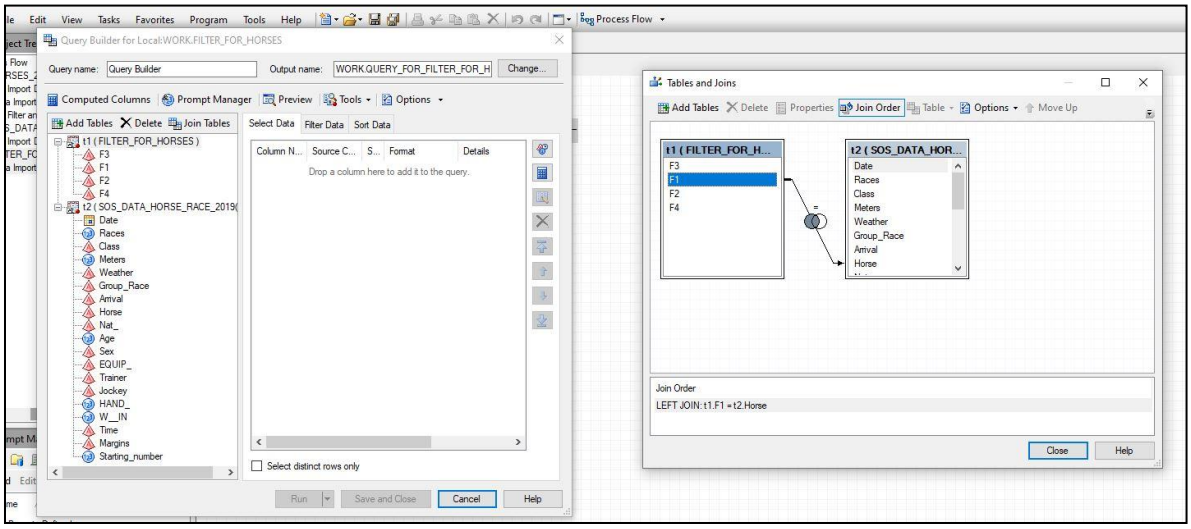
Στιγμιότυπο από τα αποτελέσματα ιπποδρομίας. Τα αποτελέσματα ερμηνεύονται από τον ισοζυγιστή και βάσει αυτών υπολογίζει τα νέα ratings για τους ίππους που διαγωνίστηκαν.

5 *Συγκέντρωση και επιλογή δεδομένων*

Προτού προχωρήσουμε στην επίλυση του προβλήματος θα πρέπει να φέρουμε τα δεδομένα στη μορφή που μπορούν να αξιοποιηθούν από το μοντέλο πρόβλεψης που θα χρησιμοποιήσουμε. Ακόμα θα πρέπει να κάνουμε έναν διαχωρισμό των δεδομένων έτσι ώστε να χρησιμοποιήσουμε μονάχα αυτά που είναι χρήσιμα για τον αλγόριθμό μας ως ανεξάρτητες μεταβλητές. Στο τέλος θα παρουσιάσουμε ακόμα πως υλοποιούμε το μοντέλο πρόβλεψης και σε ποιο σημείο εισάγουμε σε αυτό τα δεδομένα που αξιοποιήσαμε στα προηγούμενα βήματα.

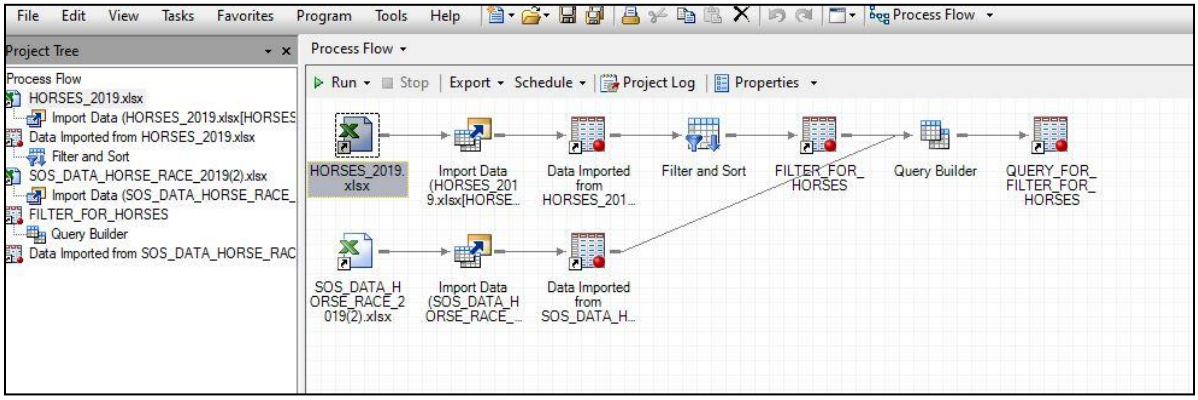
5.1 Συμψηφισμός δεδομένων

Για το συμψηφισμό των δεδομένων θα χρησιμοποιήσουμε το SAS. Το SAS είναι μια πλατφόρμα ανάλυσης δεδομένων που μας επιτρέπει το συμψηφισμό και την επεξεργασία δεδομένων από διαφορετικές γλώσσες προγραμματισμού και περιβάλλοντα. Για αρχή θα πρέπει να ¨κρατήσουμε¨ στην πορεία του προβλήματος μόνο τα άλογα τα οποία έτρεξαν σε αγώνα μέσα στο 2019. Για να γίνει αυτό θα φιλτράρουμε τη λίστα με τα ονόματα των αλόγων που πήραν Rating μέσα στο 2019. Ουσιαστικά επιλέξαμε να δημιουργήσουμε ένα query το οποίο έχει ως στοιχείο-σύνδεσμο το όνομα του αλόγου και μας δίνει έναν νέο πίνακα με τα ενεργά άλογα για την περίοδο που μας ενδιαφέρει. Στο συγκεκριμένο πρόγραμμα η υλοποίηση θα μπορούσε να γίνει και με άλλους τρόπους, όμως επιλέξαμε τον πιο γρήγορο από τους διαθέσιμους.



Οθόνη από τη δημιουργία query στο SAS

Με τη δημιουργία του project στο SAS μπορούμε να δούμε εποπτικά την μορφή της διεργασίας που γίνεται για το φιλτράρισμα αυτό των δεδομένων. Το αποτέλεσμα είναι ένα νέο φύλλο δεδομένων που έχει μόνο τα απαιτούμενα στοιχεία.



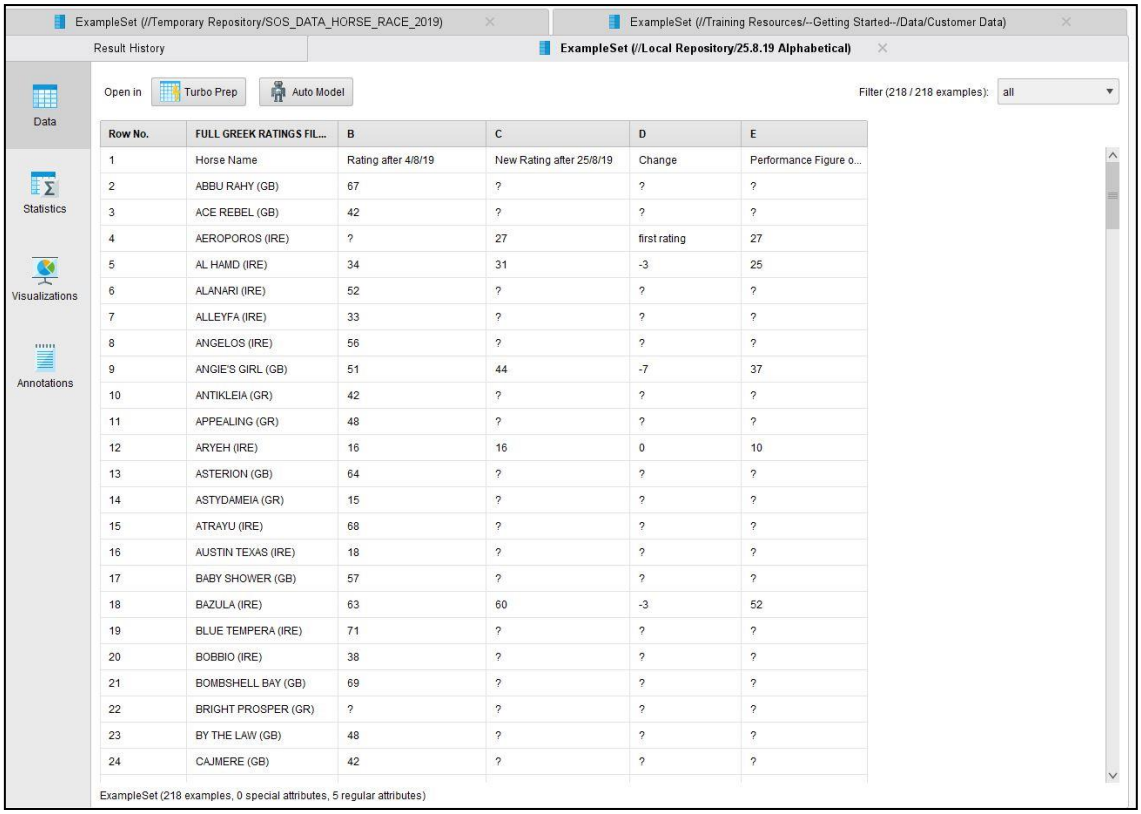
Οθόνη από τη δημιουργία project στο SAS.

Ακόμα, υπάρχει στα στοιχεία μας η πληροφορία για το rating του κάθε ίππου και πως αυτό μεταβάλλεται κατά τη διάρκεια του έτους. Στο σημείο αυτό θα πρέπει να γίνει η συγχώνευση των Rating για κάθε ίππο σε κάθε ημερομηνία, έτσι ώστε η πληροφορία να είναι συγκεντρωμένη σε ένα σημείο.

5.2 Υπολογισμός των ανεξάρτητων μεταβλητών

Με όλα τα στοιχεία συγκεντρωμένα σε ένα φύλλο δεδομένων θα χρησιμοποιήσουμε το RapidMiner για τον προσδιορισμό των ανεξάρτητων μεταβλητών.

Το RapidMiner χρησιμοποιεί ένα σύνολο από χειριστές (operators) με σκοπό την αξιοποίηση ή/και το μετασχηματισμό των δεδομένων. Τα δομικά στοιχεία, ομαδοποιημένα κατά λειτουργία, χρησιμοποιούνται για τη δημιουργία διαδικασιών RapidMiner. Ένας χειριστής έχει θύρες εισόδου και εξόδου(input/output). Η ενέργεια που εκτελείται στην είσοδο τελικά οδηγεί σε αυτό που παρέχεται στην έξοδο. Για τον προσδιορισμό των στοιχείων που θα είναι οι ανεξάρτητες μεταβλητές θα πρέπει για αρχή να εισάγουμε τα δεδομένα στο RapidMiner.



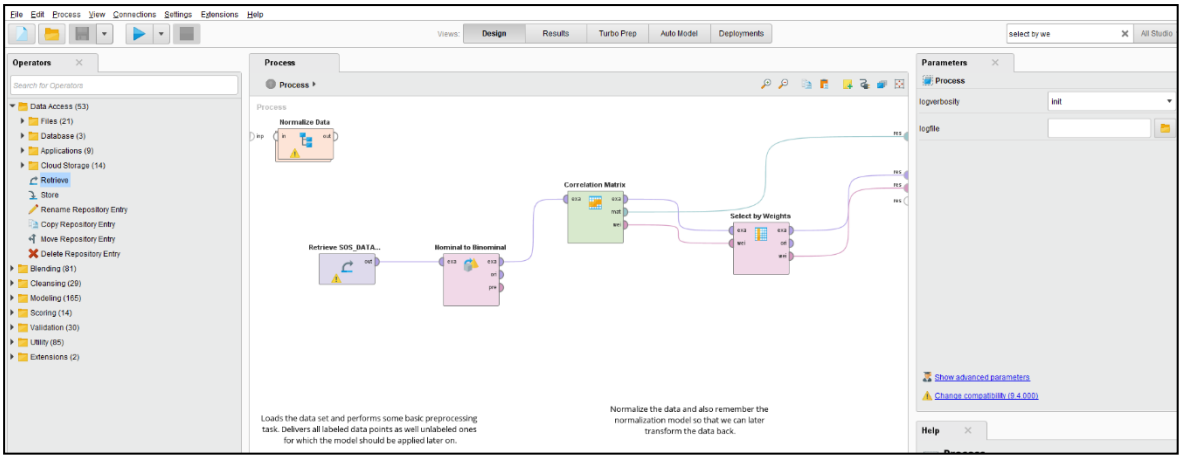
Row No.	FULL GREEK RATINGS FIL...	B	C	D	E
1	Horse Name	Rating after 4/8/19	New Rating after 25/8/19	Change	Performance Figure o...
2	ABBU RAHY (GB)	67	?	?	?
3	ACE REBEL (GB)	42	?	?	?
4	AEROPOROS (IRE)	?	27	first rating	27
5	AL HAMD (IRE)	34	31	-3	25
6	ALANARI (IRE)	52	?	?	?
7	ALLEYFA (IRE)	33	?	?	?
8	ANGELOS (IRE)	56	?	?	?
9	ANGIE'S GIRL (GB)	51	44	-7	37
10	ANTIKLEIA (GR)	42	?	?	?
11	APPEALING (GR)	48	?	?	?
12	ARYEH (IRE)	16	16	0	10
13	ASTERION (GB)	64	?	?	?
14	ASTYDAMEIA (GR)	15	?	?	?
15	ATRAYU (IRE)	68	?	?	?
16	AUSTIN TEXAS (IRE)	18	?	?	?
17	BABY SHOWER (GB)	57	?	?	?
18	BAZULA (IRE)	63	60	-3	52
19	BLUE TEMPERA (IRE)	71	?	?	?
20	BOBBIO (IRE)	38	?	?	?
21	BOMBSHELL BAY (GB)	69	?	?	?
22	BRIGHT PROSPER (GR)	?	?	?	?
23	BY THE LAW (GB)	48	?	?	?
24	CAJMERE (GB)	42	?	?	?

Οθόνη από τα δεδομένα στο περιβάλλον του RapidMiner

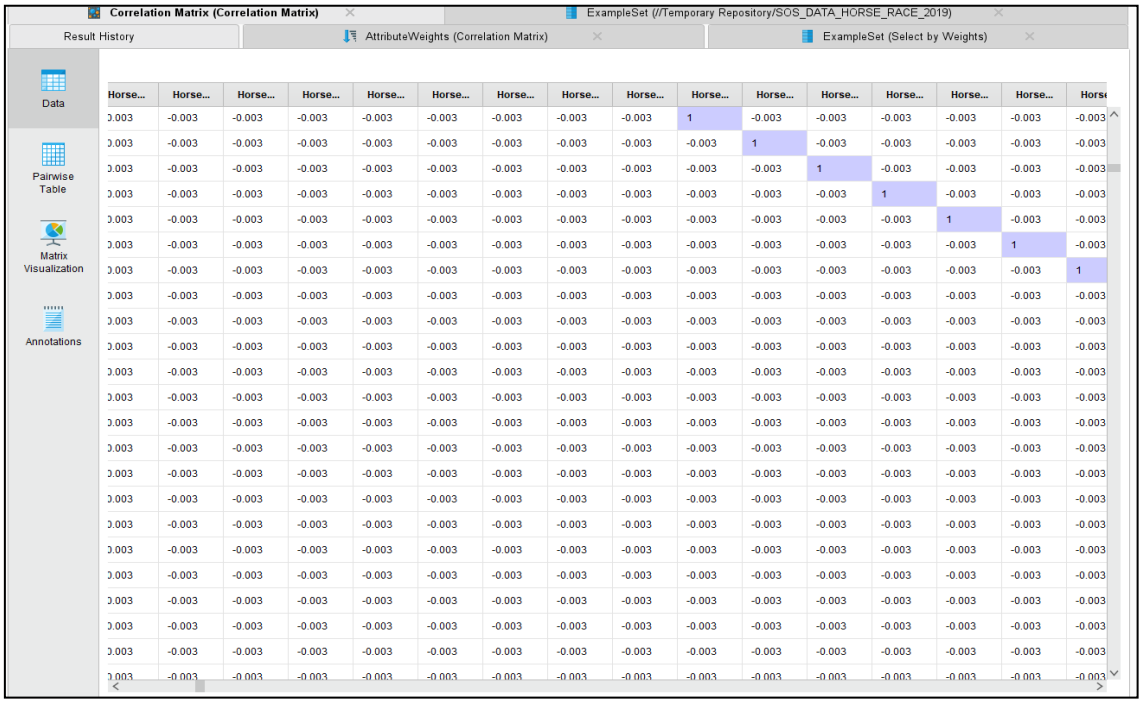
Από τα εργαλεία του RapidMiner θα χρησιμοποιήσουμε τον Πίνακα Συσχέτισης. Αυτός θα μας δώσει τα στοιχεία για το πλήθος των ανεξάρτητων μεταβλητών. Μια συσχέτιση είναι ένας αριθμός μεταξύ -1 και +1 που μετρά το βαθμό συσχέτισης μεταξύ δύο χαρακτηριστικών (attributes) (έστω x και ψ). Μια θετική τιμή για τον συσχετισμό συνεπάγεται μια θετική συσχέτιση. Σε αυτή την περίπτωση μεγάλες τιμές του x τείνουν να συνδέονται με μεγάλες τιμές του ψ και μικρές τιμές του x τείνουν να συσχετίζονται με μικρές τιμές του ψ. Μια αρνητική τιμή για τον συσχετισμό συνεπάγεται αρνητική ή αντίστροφη συσχέτιση(μπορεί να είναι αντιστρόφως ανάλογα ποσά). Σε αυτή την περίπτωση οι μεγάλες τιμές του x τείνουν να συσχετίζονται με μικρές τιμές του ψ και αντιστρόφως.

Ο πίνακας συσχέτισης μπορεί να χρησιμοποιηθεί για τη δημιουργία ενός πίνακα που δείχνει συσχετισμούς όλων των ιδιοτήτων της εισόδου. Το RapidMiner επιστρέφει και ένα διάνυσμα βαρύτητας του κάθε χαρακτηριστικού. Τα συσχετισμένα χαρακτηριστικά απομακρύνονται συνήθως επειδή είναι παρόμοια στη συμπεριφορά και έχουν μικρή μόνο επιρροή κατά τον υπολογισμό των προβλέψεων. Μπορεί επίσης να παρεμποδίσουν το χρόνο εκτέλεσης και τη χρήση της μνήμης.

Για τη διαδικασία υπολογισμού των ανεξάρτητων μεταβλητών στήσαμε την διαδικασία στο πρόγραμμα.



Διαδικασία υπολογισμού του πίνακα συσχέτισης στο περιβάλλον του RapidMiner



Οθόνη από τον πίνακα συσχέτισης του RapidMiner

Το αποτέλεσμα του πίνακα συσχέτισης μας δείχνει πως δεν υπάρχει σημαντική συσχέτιση για κάποια από τα στοιχεία εισόδου. Αυτό σημαίνει πως μπορούν όλα να ληφθούν ως ανεξάρτητες μεταβλητές του προβλήματος.

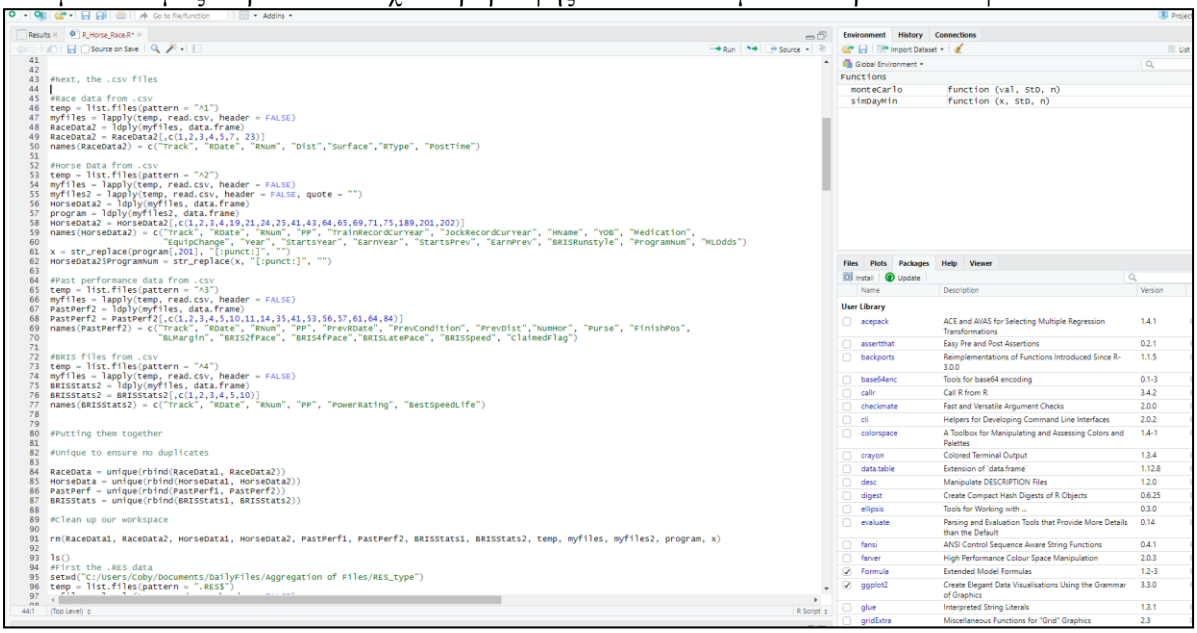
Επομένως οι μεταβλητές για το πρόβλημα της πρόβλεψης είναι :

Ανεξάρτητη Μεταβλητή	***Περιγραφή***
1	Μήκος Διαδρομής
2	Φύλο ίππου
3	Ηλικία ίππου
4	Αναβάτης ίππου
5	Προπονητής ίππου
6	Rating ίππου

5.3 Υλοποίηση του μοντέλου πρόβλεψης με κώδικα

Εκκινώντας από τη βάση που θέσαμε παραπάνω ξεκινάμε την υλοποίηση ενός αλγορίθμου πρόβλεψης των αγώνων σε γλώσσα R. Θα πρέπει να γίνει μετατροπή των αρχείων σε μορφή που να μπορεί να γίνουν χρήσιμα για το πρόγραμμα.

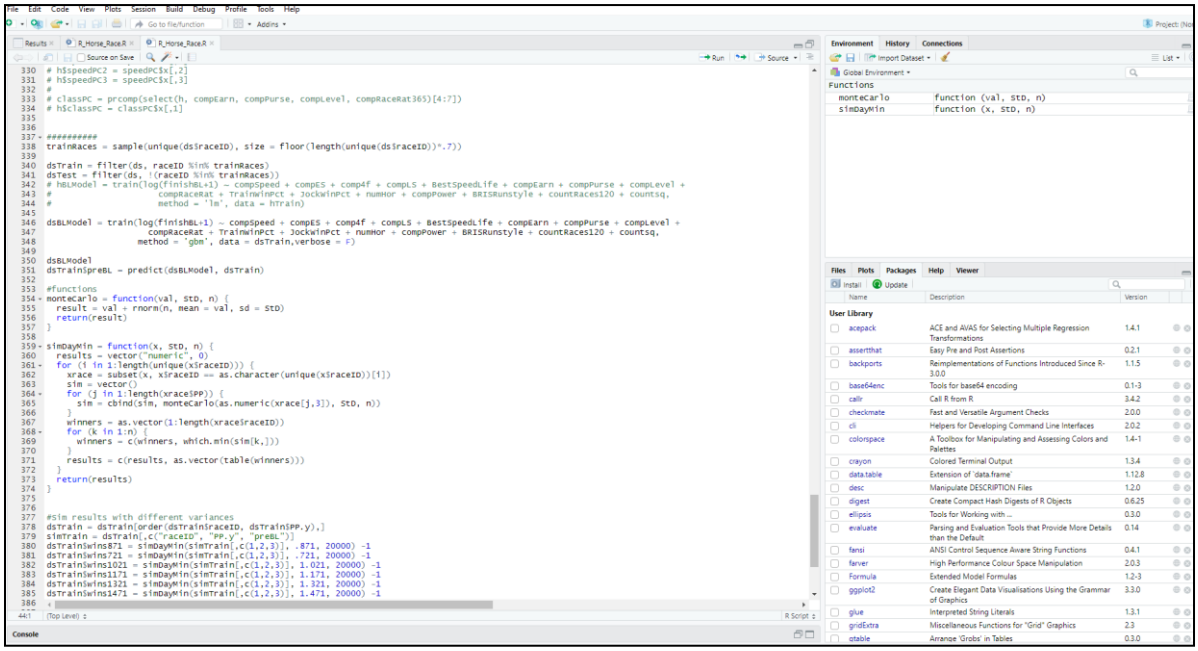
Μετά την εισαγωγή των δεδομένων θα χρησιμοποιήσουμε κάποια κομμάτια κώδικα για να ορίσουμε το ανταγωνιστικό στοιχείο εντός της κούρσας. Στόχος είναι να συμπεριληφθούν στο μοντέλο μας τα βασικά στοιχεία πρόβλεψης που αναλύσαμε στα παραπάνω κεφάλαια.



Οθόνη από το πρόγραμμα πρόβλεψης σε γλώσσα R

Ο στόχος του προγράμματος είναι:

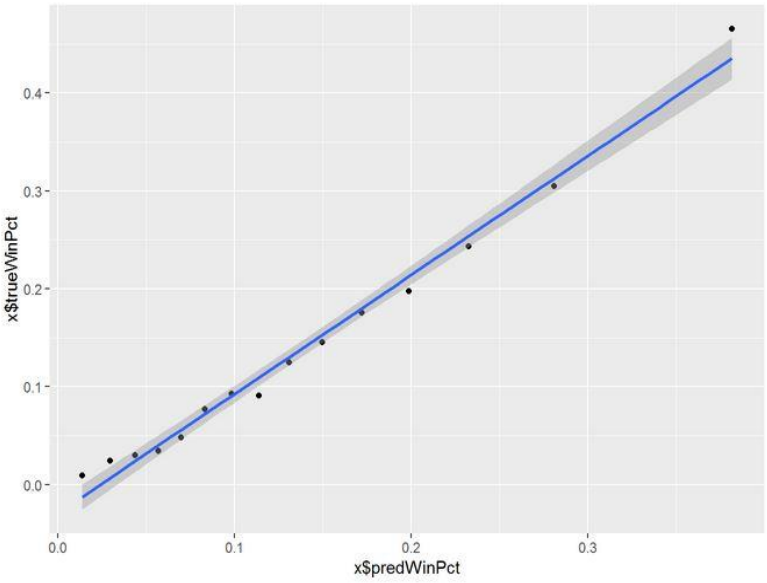
- Να προετοιμάσει τα δεδομένα σε μορφή που τα διαβάζει η R.
- Να «διαβάσει» τα αποτελέσματα των παρελθόντων αγώνων.
- Να κάνει υπολογισμούς για τον ανταγωνισμό εντός του γεγονότος.
- Να κάνει παλινδρόμηση και να υπολογίσει τη σειρά των νικητών.
- Χωρισμό του μοντέλου σε trainset και testset .Για το trainset χρησιμοποιούμε τις πρώτες κούρσες από τα δεδομένα μας , ενώ για το testset τις υπόλοιπες. Ο χωρισμός γίνεται 10 % trainset , 90 % test set
- Αξιολόγηση του μοντέλου με βάσει τις αποδόσεις της αγοράς και την κερδοφορία. Αναγράφεται αναλυτικά στο παρακάτω κεφάλαιο.



Οθόνη από το πρόγραμμα πρόβλεψης σε γλώσσα R

```
x = group_by(dsTrain, group1771)
x = summarise(x, trueWinPct = sum(FinishPos == 1)/length(raceID), predWinPct = mean(wins1771)/20000)

qplot(x$predWinPct, x$trueWinPct) + geom_smooth(method = 'lm')
```



Οθόνη από το αποτέλεσμα της παλινδρόμησης στην R . Όσο τα σημεία φτιάχνουν μια καλύτερη ευθεία (μικρή γκρίζα περιοχή) τόσο το μοντέλο βρίσκει καλύτερη γραμμική συσχέτιση για τα διάφορα trainset και test set.

6 *Αξιολόγηση του μοντέλου πρόβλεψης*

Παρά την ανάπτυξη και την υλοποίηση του μοντέλου πρόβλεψης θα πρέπει να δοθεί απάντηση στο ερώτημα: Πόσο σωστές είναι οι προβλέψεις του μοντέλου; Παρακάτω θα δούμε αναλυτικά τις παραμέτρους βάσει των οποίων αξιολογείται το μοντέλο πρόβλεψης που παρουσιάσαμε παραπάνω.

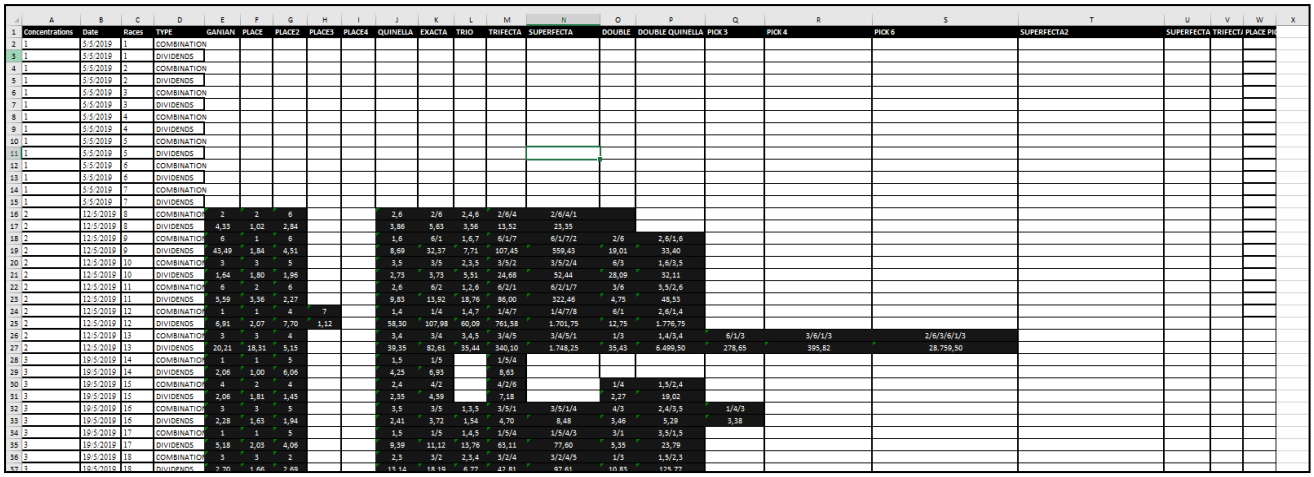
6.1 Παράμετροι αξιολόγησης

Προφανώς δε μπορούμε να περιμένουμε απόλυτη επιτυχία του οποιουδήποτε μοντέλου πρόβλεψης. Επίσης έχει διαφορετική αξία η πρόβλεψη του νικητή μιας ιπποδρομίας όπου όλοι οι διαγωνιζόμενοι είναι στο ίδιο επίπεδο και άλλη αξία όταν προβλέπεται ένα δυνατό φαβορί και αυτό επικρατεί.

Θα ήταν λάθος σε πρώτη φάση να κρίνουμε με το ποσοστό επιτυχίας του μοντέλου στις προβλέψεις νίκης. Ένα κριτήριο που θα μπορούσε να είναι δείκτης για το μοντέλο μας είναι στο κατά πόσο θα μπορούσε να κερδίσει την αγορά στοιχημάτων (beat the market). Αυτό αποτελεί το κριτήριο για όλα τα μοντέλα και ο τρόπος που αξιολογούνται. Γενικότερα εάν οι προβλέψεις μας μπορούν να αποφέρουν κέρδος μακροπρόθεσμα τότε το μοντέλο μπορεί να κριθεί ως επιτυχές και το χρησιμοποιούμε. Εάν δεν αποφέρει κέρδος μπορούμε εκ των υστέρων να δούμε τους λόγους που αυτό συμβαίνει και να βελτιώσουμε κάποιες παραμέτρους που αποτρέπουν την κερδοφορία.

6.2 Σύστημα αξιολόγησης

Το ερώτημα λοιπόν μετασχηματίζεται ως εξής: Μπορεί το μοντέλο μου να είναι κερδοφόρο;
Για να μπορέσουμε να αποφανθούμε κατά πόσο το μοντέλο μας μπορεί να παράγει κέρδος θα πρέπει να αντιπαραβάλλουμε τις προβλέψεις μας με τις αποδόσεις στις οποίες άνοιξε η αγορά (opening prices)



Concentrations	Date	Races	TYPE	GANIAN	PLACE	PLACE2	PLACE3	PLACE4	QUINELLA	EXACTA	TRIO	TRIFECTA	SUPERFECTA	DOUBLE	DOUBLE QUINELLA	PICK 3	PICK 4	PICK 6	SUPERFECTA2	SUPERFECTA	TRIFECTA	PLACE PI
1	5/5/2019	1	COMBINATION																			
1	5/5/2019	1	DIVIDENDS																			
1	5/5/2019	2	COMBINATION																			
1	5/5/2019	2	DIVIDENDS																			
1	5/5/2019	3	COMBINATION																			
1	5/5/2019	3	DIVIDENDS																			
1	5/5/2019	4	COMBINATION																			
1	5/5/2019	4	DIVIDENDS																			
1	5/5/2019	5	COMBINATION																			
1	5/5/2019	5	DIVIDENDS																			
1	5/5/2019	6	COMBINATION																			
1	5/5/2019	6	DIVIDENDS																			
1	5/5/2019	7	COMBINATION																			
1	5/5/2019	7	DIVIDENDS																			
2	12/5/2019	8	COMBINATION	2	2	6			2,6	2/6	2,4,6	2/6/4	2/6/4/1									
2	12/5/2019	8	DIVIDENDS	4,33	1,02	2,84			3,86	5,63	3,56	13,52	23,35									
2	12/5/2019	9	COMBINATION	6	1	6			1,6	6/1	1,6,7	6/1/7	6/1/7/2	2/6	2,6/1,6							
2	12/5/2019	9	DIVIDENDS	43,49	1,84	4,51			8,69	32,37	7,71	107,45	559,43	19,01	33,40							
2	12/5/2019	10	COMBINATION	3	3	5			3,5	3/5	2,3,5	3/5/2	3/5/2/4	6/3	1,6/3,5							
2	12/5/2019	10	DIVIDENDS	1,64	1,80	1,96			2,73	3,73	5,51	24,68	52,44	28,09	32,11							
2	12/5/2019	11	COMBINATION	6	2	6			2,6	6/2	1,2,6	6/2/1	6/2/1/7	3/6	3,5/2,6							
2	12/5/2019	11	DIVIDENDS	5,59	3,36	2,27			9,83	13,92	18,76	86,00	322,48	4,75	48,53							
2	12/5/2019	12	COMBINATION	1	1	4	7		1,4	1/4	1,4,7	1/4/7	1/4/7/8	6/1	2,6/1,4							
2	12/5/2019	12	DIVIDENDS	6,91	2,07	7,70	1,12		58,30	107,98	60,09	761,58	1.701,75	12,75	1.776,75							
2	12/5/2019	13	COMBINATION	3	3	4			3,4	3/4	3,4,5	3/4/5	3/4/5/1	1/3	1,4/3,4	6/1/3	3/6/1/3	2/6/3/6/1/3				
2	12/5/2019	13	DIVIDENDS	20,21	18,31	5,15			39,35	82,61	35,44	340,10	1.748,25	35,43	6.499,50	278,65	395,82	28.759,50				
3	19/5/2019	14	COMBINATION	1	1	5			1,5	1/5		1/5/4										
3	19/5/2019	14	DIVIDENDS	2,06	1,00	6,06			4,25	6,93		8,63										
3	19/5/2019	15	COMBINATION	4	2	4			2,4	4/2		4/2/6		1/4	1,5/2,4							
3	19/5/2019	15	DIVIDENDS	2,06	1,81	1,45			2,35	4,59		7,18		2,27	19,02							
3	19/5/2019	16	COMBINATION	3	3	5			3,5	3/5	1,3,5	3/5/1	3/5/1/4	4/3	2,4/3,5	1/4/3						
3	19/5/2019	16	DIVIDENDS	2,28	1,63	1,94			2,41	3,72	1,54	4,70	8,48	3,46	5,29	3,38						
3	19/5/2019	17	COMBINATION	1	1	5			1,5	1/5	1,4,5	1/5/4	1/5/4/3	3/1	3,5/1,5							
3	19/5/2019	17	DIVIDENDS	5,18	2,03	4,06			9,39	11,12	13,76	63,11	77,60	5,35	23,79							
3	19/5/2019	18	COMBINATION	3	3	2			2,3	3/2	2,3,4	3/2/4	3/2/4/5	1/3	1,5/2,3							
3	19/5/2019	18	DIVIDENDS	2,20	1,66	2,48			13,16	18,19	6,77	47,81	97,61	10,83	124,77							

Συγκεκριμένα για τη μεγιστοποίηση του κέρδους δεν αρκεί απλά να ποντάρουμε στον εκάστοτε νικητή της ιπποδρομίας αλλά να εφαρμόσουμε μια συγκεκριμένη στρατηγική στοιχημάτων, όπως θα αναλύσουμε παρακάτω.

Ακόμα, για να αποφανθούμε κατά πόσο είναι κερδοφόρο το μοντέλο μας θα πρέπει να δούμε τις χρηματοροές μας μετά από τα στοιχήματα σαν να επρόκειτο για κάποια επένδυση.

6.3 Οργάνωση του συστήματος αξιολόγησης

Για να ελέγξουμε την κερδοφορία μας θα εφαρμόσουμε στοιχηματισμό σύμφωνα με το κριτήριο Kelly. Αποτελεί ένα είδος στρατηγικής τυχερών παιχνιδιών που δημιουργήθηκε από τον J.L. Kelly Jr. το 1956 για να βοηθήσει στη μεγιστοποίηση του κέρδους μιας σειράς στοιχημάτων. Είναι γνωστό με πολλά ονόματα όπως ο κανόνας Kelly ή το στοίχημα Kelly, εκτός από το όνομα της στρατηγικής.

Με λίγα λόγια, η στρατηγική Kelly είναι ένας τρόπος για τους παίκτες (ή τους επενδυτές στη χρηματιστηριακή αγορά) να εξετάσουν λεπτομερώς τους κινδύνους και τις ανταμοιβές κάθε στοιχήματος ή επένδυσης. Ο συγκεκριμένος τύπος έχει ως εξής:

$$f = \frac{bp - q}{b} = \frac{p(b+1) - 1}{b}$$

Το f αντιπροσωπεύει το ποσό του στοιχήματος ή του τρέχοντος χαρτοφυλακίου. Το γράμμα b αντιπροσωπεύει τις αποδόσεις που εισπράχθηκαν στο στοίχημα ή πόσα χρήματα θα μπορούσατε να κερδίσετε εκτός από τα χρήματα που έχετε ποντάρει. Η πιθανότητα νίκης υποδεικνύεται από το γράμμα p, και q είναι η πιθανότητα να χάσει. Φυσικά αυτό σημαίνει πως σε κάθε ιπποδρομία θα ποντάρουμε κάποια χρήματα σε παραπάνω από ένα άλογα με τη λογική ότι πάντα υπάρχει η πιθανότητα (προγνωστικά) να είναι απίθανο να προκρίνουμε έναν και μοναδικό νικητή σε κάθε ιπποδρομία.

Ο τύπος αυτός πρέπει να ακολουθηθεί κατά γράμμα για να λειτουργήσει η στρατηγική Kelly. Πρόκειται για ένα σύστημα σχεδιασμένο με μακροπρόθεσμη κερδοφορία, σε αντίθεση με τα άμεσα ή βραχυπρόθεσμα κέρδη.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Η καλύτερη επιλογή είναι συνήθως να επενδύετε μικρότερο ποσό από το πλήρες ποσό Kelly. Ακόμα κι αν αυτό μειώνει το ρυθμό ανάπτυξης κέρδους, μειώνει επίσης τον κίνδυνο απώλειας υψηλότερων χρηματικών ποσών γρήγορα. Χρησιμοποιώντας τον παραπάνω τύπο, μπορεί να προσδιορισθεί μια τάξη μεγέθους του ποσού που μπορεί να επενδυθεί / πονταριστεί σε μια δεδομένη κατάσταση. Η στρατηγική Kelly βοηθά πολλούς επαγγελματίες παίκτες να υπολογίζουν πόσο πρέπει να ποντάρουν σε οποιοδήποτε συγκεκριμένο στοίχημα, ακόμα κι αν οι πιθανότητες είναι υπέρ τους. Αυτή η μέθοδος τους βοηθά ακόμα και αν κερδίζουν, επειδή τους επιτρέπει να συνεχίσουν να ποντάρουν ακόμα και όταν βρίσκονται σε καλό ή κακό σερί.

Φυσικά, το παιχνίδι δεν έχει ποτέ εγγύηση 100% ότι θα βγει κερδοφόρο και αυτό θα πρέπει να έχει κανείς κατά νου, ακόμη και όταν χρησιμοποιούνται εργαλεία όπως το στοίχημα Kelly για τη μεγιστοποίηση των κερδών. Παρόλα αυτά αποτελεί έναν καλό τρόπο για την υιοθέτηση στοιχηματικών τακτικών για ερευνητικούς σκοπούς – όπως και στην παρούσα εργασία- και να εξασφαλιστεί μια γενικότερα θετική εμπειρία όταν μέσα από τις επενδύσεις.

Όσον αφορά την οπτική μας ως επένδυσης , η βασική ιδέα είναι να εκφραστούν τα πρόσθετα χρήματα ή την αξία που έχει βρεθεί - το όφελος ή την απόδοση που έγινε - ως ποσοστό της αρχικής επένδυσης. Το Rate of Investment (ή για συντομία ROI) είναι ένας πολύ χρήσιμος τύπος που έχει ως εξής:

ROI = (Επιστροφή / αρχική επένδυση) x 100

Πολλαπλασιάζοντας με το 100 ο λόγος μετατρέπεται σε ποσοστό. Για παράδειγμα, αγοράζει κάποιος μια μικρή επιχείρηση για $ 200.000. Μέσα από σκληρή δουλειά, αναπτύσσει την επιχείρηση και την πουλάει για $ 300.000. Η απόδοση είναι η τελική τιμή πώλησης των $ 300.000 μείον την τιμή αγοράς, την επένδυση, ύψους $ 200.000.

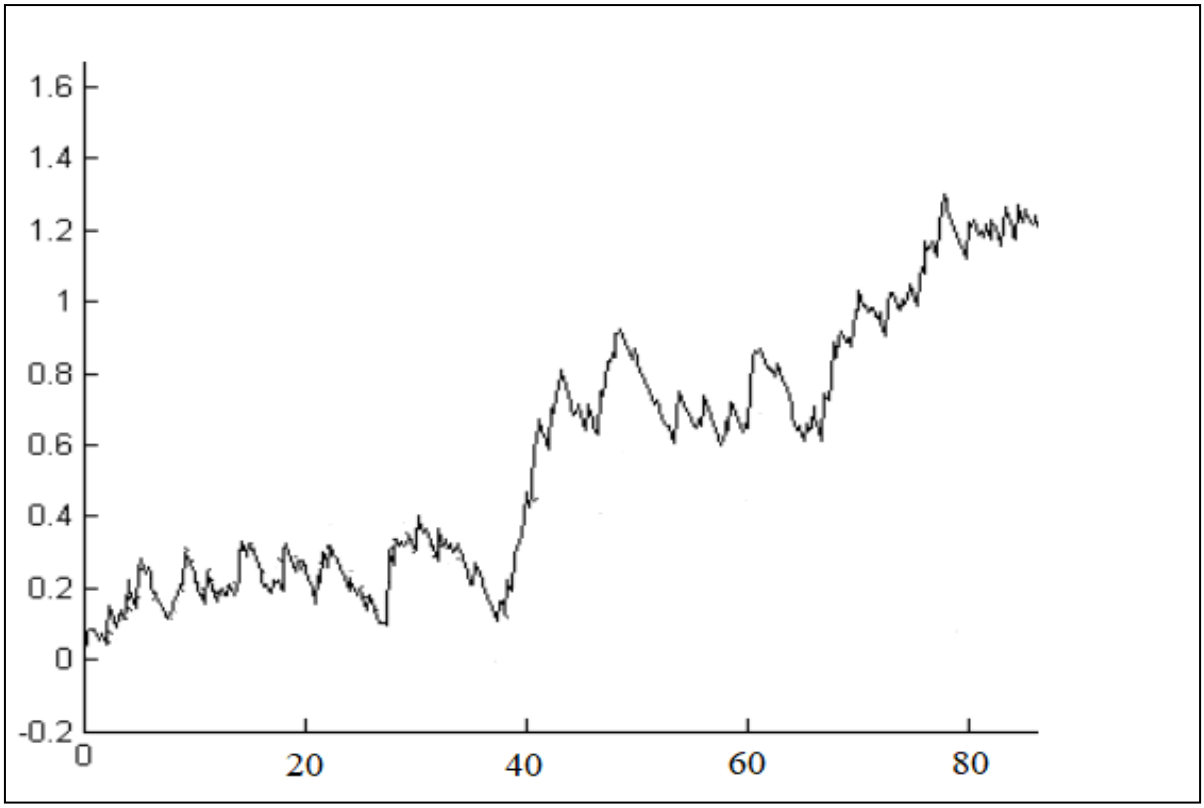
Το κέρδος είναι $ 100.000. Εάν χωρίσει κανείς το κέρδος από την επένδυση φαίνεται ότι κερδίζει 0,50. Πολλαπλασιάζοντας με το 100 βρίσκουμε πως η απόδοση της επένδυσης (ROI) ήταν 50%.

6.4 Αποτελέσματα

Μετά από την εισαγωγή στις έννοιες που πραγματοποιήσαμε παραπάνω θα κάνουμε ένα διάγραμμα που θα δείχνει τη σχέση του ROI της επένδυσής μας με τον αριθμό των προβλέψεων (στοιχημάτων) στις ιπποδρομίες του 2019. Η επένδυση αποδίδει το ποσό που στοιχηματίζουμε επί την απόδοση νίκης του αλόγου, όπως αυτή παρέχεται στο ιστορικό από το site του ΟΠΑΠ για τις ιπποδρομίες που διεξήχθησαν το 2019. Σε κάθε περίπτωση το ποσό στοιχήματος δεν είναι το ίδιο σε κάθε γεγονός και δεν στοιχηματίζετέ ο ίδιος αριθμός νικητών κάθε φορά.

Έστω ότι ξεκινάμε με αρχικό κεφάλαιο 5.000 € και ξεκινάμε έχοντας μπροστά μας 100 αγώνες. Με το τελευταίο μέρος του αλγορίθμου ορίζουμε ως εξαρτημένη μεταβλητή (άξονας x) τον αριθμό των αγώνων και ως ανεξάρτητη μεταβλητή (άξονας y) το ROI της επένδυσης.

ΣΗΜΕΙΩΣΗ: Λαμβάνουμε ως αποδόσεις στις οποίες τοποθετούνται στοιχήματα τις αποδόσεις με τις οποίες ανοίγει η αγορά . Αυτές είναι διαθέσιμες στο site των ιπποδρομιών



Διάγραμμα απόδοσης επένδυσης (ROI) σε συνάρτηση με τον αριθμό αγώνων στους οποίους τοποθετήθηκαν στοιχήματα σύμφωνα με το κριτήριο Kelly και σύμφωνα με τις προτάσεις του μοντέλου για τις ιπποδρομίες του 2019.

6.5 Σύνοψη συμπερασμάτων αξιολόγησης

Παρά το γεγονός ότι έχουμε δώσει έναν αρκετά μεγάλο αριθμό αγώνων και έχουμε υιοθετήσει μια από τις πιο αποτελεσματικές τεχνικές στοιχηματισμού, το ROI της επένδυσης δεν ξεπερνάει το 1.4 %. Επομένως μπορούμε να αποφανθούμε για το εάν το εργαλείο πρόβλεψης είναι καλό ή όχι;

Το θετικό είναι πως μπορούμε και έχουμε μεγαλύτερες θετικές από ότι αρνητικές χρηματοροές . Αυτό σημαίνει πως σε βάθος χρόνου μπορούμε, έστω και οριακά, να δούμε το αρχικό μας κεφάλαιο να αυξάνεται. Το πλήθος των αγώνων στους οποίους εφαρμόσαμε το μοντέλο μας δεν κρίνεται ως πολύ μεγάλος, όμως δείχνει να φέρνει καλά αποτελέσματα. Βέβαια υπάρχουν κάποιες παραδοχές στις οποίες έχουμε προβεί για την δημιουργία του γραφήματος. Η βασικότερη από αυτές είναι πως οι αποδόσεις που παίρνουμε είναι οι αρχικές (starting prices) κάτι που δεν είναι απόλυτα σωστό, καθώς η αγορά στοιχημάτων είναι ένα πολύ δυναμικό περιβάλλον και οι αποδόσεις μεταβάλλονται πολύ γρήγορα. Μόλις απομένουν λίγα λεπτά για την έναρξη της κούρσας οι αποδόσεις αλλάζουν κάθε λίγα δευτερόλεπτα! Τέλος η άλλη παραδοχή που έχουμε κάνει είναι πως το μοντέλο κρίνεται με το δείκτη ROI και όχι με κάποιον άλλο , όπως πχ το NPV ή το IRR που θα μπορούσαν να δίνουν πιο ακριβή αποτελέσματα.

7 Τεχνικές λεπτομέρειες

Η συγγραφή της εργασίας έγινε με τη βοήθεια των εργαλείων R Studio, SAS & Rapidminer. Παρακάτω θα παρουσιάσουμε κάποια θέματα που παρουσιάζουν ενδιαφέρον όσον αφορά την πορεία υλοποίησης κάποιων σημείων της εργασίας.

7.1 Λεπτομέρειες υλοποίησης

Σε πρώτη φάση χρησιμοποιήσαμε το RapidMiner για να είμαστε σίγουροι ότι δίνουμε στο μοντέλο μας ανεξάρτητες μεταβλητές.

7.1.1 Πίνακας συσχέτισης

Το RapidMiner είναι ένα ολοκληρωμένο πακέτο Data Analysis που το χρησιμοποιούμε γιατί μπορεί να χειριστεί πολλές μορφές αρχείων. Ακόμα έχει έτοιμες λειτουργίες για λογικούς και μαθηματικούς τελεστές. Ένας από αυτούς είναι ο πίνακας συσχέτισης. Έστω δύο μεταβλητές χ και ψ. Όπως αναφέρθηκε και παραπάνω μια θετική τιμή για τη συσχέτιση συνεπάγεται μια θετική συσχέτιση. Αν η τιμή χ και η τιμή ψ ήταν και οι δύο κάτω από το μέσο όρο, τότε το παραπάνω προϊόν θα είχε δύο αρνητικούς αριθμούς, κάτι που θα ήταν επίσης θετικό.

Επομένως, ένας θετικός συσχετισμός είναι απόδειξη μιας γενικής τάσης ότι μεγάλες τιμές του χ σχετίζονται με μεγάλες τιμές του ψ και μικρές τιμές του χ συνδυάζονται με μικρές τιμές του ψ. Στην στατιστική επιστήμη είναι κάτι που χρησιμοποιείται σε μεγάλο βαθμό και το υιοθετήσαμε και στην προετοιμασία των δεδομένων μας , πιο πολύ ως μια μέθοδο επικύρωσης για τις μεταβλητές μας .



Χειριστής (operator) Πίνακας Συσχέτισης στο RapidMiner

7.1.2 Κώδικας r

Στο επίκεντρο του προβλήματος είναι η δημιουργία του κώδικα που θα κάνει τις προβλέψεις . Το σημαντικό κομμάτι για κάθε πρόγραμμα όπως και αυτό είναι να μπορέσει να έχει αρκετά δεδομένα έτσι ώστε να κάνει εξάσκηση (trainset) αλλά και να παράγει αποτέλεσμα.

```
dsTrain$preBL = predict(dsBLModel, dsTrain)
```

```
#functions
monteCarlo = function(val, StD, n) {
  result = val + rnorm(n, mean = val, sd = StD)
  return(result)
}

simDayMin = function(x, StD, n) {
  results = vector("numeric", 0)
  for (i in 1:length(unique(x$raceID))) {
    xrace = subset(x, x$raceID == as.character(unique(x$raceID))[i])
    sim = vector()
    for (j in 1:length(xrace$PP)) {
      sim = cbind(sim, monteCarlo(as.numeric(xrace[j,3]), StD, n))
    }
    winners = as.vector(1:length(xrace$raceID))
    for (k in 1:n) {
      winners = c(winners, which.min(sim[k,]))
    }
    results = c(results, as.vector(table(winners)))
  }
  return(results)
}


#Sim results with different variances
dsTrain = dsTrain[order(dsTrain$raceID, dsTrain$PP.y),]
simTrain = dsTrain[,c("raceID", "PP.y", "preBL")]
dsTrain$wins871 = simDayMin(simTrain[,c(1,2,3)], .871, 20000) -1
dsTrain$wins721 = simDayMin(simTrain[,c(1,2,3)], .721, 20000) -1
dsTrain$wins1021 = simDayMin(simTrain[,c(1,2,3)], 1.021, 20000) -1
dsTrain$wins1171 = simDayMin(simTrain[,c(1,2,3)], 1.171, 20000) -1
dsTrain$wins1321 = simDayMin(simTrain[,c(1,2,3)], 1.321, 20000) -1
dsTrain$wins1471 = simDayMin(simTrain[,c(1,2,3)], 1.471, 20000) -1
dsTrain$wins1621 = simDayMin(simTrain[,c(1,2,3)], 1.621, 20000) -1
dsTrain$wins1771 = simDayMin(simTrain[,c(1,2,3)], 1.771, 20000) -1


#group horses by approximate percentage of winning
#I use 4 groups with about equal numbers of horses
library(Hmisc)
dsTrain$group721 = cut2(dsTrain$wins721, g = 4)
dsTrain$group871 = cut2(dsTrain$wins871, g = 4)
dsTrain$group1021 = cut2(dsTrain$wins1021, g = 4)
dsTrain$group1171 = cut2(dsTrain$wins1171, g = 4)
dsTrain$group1321 = cut2(dsTrain$wins1321, g = 4)
dsTrain$group1471 = cut2(dsTrain$wins1471, g = 4)
dsTrain$group1621 = cut2(dsTrain$wins1621, g = 4)
dsTrain$group1771 = cut2(dsTrain$wins1771, g = 4)
################################################################################
######################
################################################################################
######################
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Χρησιμοποιούμε και την προσομοίωση monte carlo, ώστε να πάρουμε καλύτερα αποτελέσματα για τον trainset μας. Σκοπός του προγράμματος είναι να υπάρξει κερδοφορία απέναντι στις αποδόσεις με τις οποίες ανοίγει η αγορά. Το κατά πόσο θα είναι τα αποτελέσματα επιθυμητά δε μπορούμε να το γνωρίζουμε εκ των προτέρων, παρά μόνο εάν ζητήσουμε από το πρόγραμμα το διάγραμμα της αξιολόγησης στο testet.

7.2 Πλατφόρμες και προγραμματιστικά εργαλεία

Για τις απαιτήσεις της διπλωματικής χρειάστηκε η εγκατάσταση σε λειτουργικό σύστημα Windows 10 των παρακάτω εργαλείων.

- SAS Enterprise Guide 7.1 32-bit, σύμφωνα με τις οδηγίες και το manual του μεταπτυχιακού προγράμματος
- RapidMiner Studio version 9.6 διαθέσιμο σε free trial version από το σύνδεσμο : https://rapidminer.com/get-started/
- RStudio Version 1.2.5033 και R version 3.6.3 -- "Holding the Windsock"

  Κάποια libraries που χρειάστηκαν για ιδιαίτερους μαθηματικούς υπολογισμούς και δημιουργία γραφημάτων κατέβηκαν και εγκαταστάθηκαν από τα αντίστοιχα repositories για την έκδοση της R που χρησιμοποιήσαμε
- Microsoft Word 2019 και Power Point 2019 για την έκθεση και την παρουσίαση

8 *Κύρια Στοιχεία της Πρακτικής Άσκησης*

8.1 Τομέας : Περιγραφή του τμήματος (ή τμήματα) στο οποίο εργαστήκατε και ποια τα καθήκοντά σας.

Η πρακτική άσκηση έγινε στον Ιππόδρομο του Μαρκόπουλου, στον οποίο δραστηριοποιείται η Ιπποδρομίες Α.Ε.

Η Ιπποδρομίες Α.Ε., 100% θυγατρική εταιρεία της ΟΠΑΠ Α.Ε., μετά την πλήρωση των αιρέσεων που προβλέπονται στην από 24.04.2015 Σύμβαση Παραχώρησης μεταξύ της εταιρείας Ιπποδρομίες ΑΕ και του ΤΑΙΠΕΔ, συμπεριλαμβανομένης της κύρωσης της σύμβασης παραχώρησης μετά των παραρτημάτων της από την Ελληνική Βουλή, απέκτησε το αποκλειστικό δικαίωμα διοργάνωσης και διεξαγωγής επίγειου και διαδικτυακού αμοιβαίου ιπποδρομιακού στοιχήματος στην Ελλάδα για 20 και για 5 χρόνια αντίστοιχα.

Σύμφωνα με αυτό το δικαίωμα, η Ιπποδρομίες Α.Ε αναλαμβάνει:

α) την οργάνωση και διεξαγωγή ιπποδρομιών στην Ελλάδα,
β) τη διοργάνωση και διεξαγωγή αμοιβαίων στοιχημάτων σε σχέση με τις διεξαγόμενες ιπποδρομίες στην Ελλάδα,
γ) την ανάληψη δραστηριοτήτων συναφών προς τα ιπποδρομικά στοιχήματα σε εθνικό και διεθνές επίπεδο,
δ) τη διοργάνωση και διεξαγωγή προσθέτου αμοιβαίου στοιχήματος επί ιπποδρομιών (sweepstake) και
ε) τη διοργάνωση και διεξαγωγή αμοιβαίων στοιχημάτων επί ιπποδρομιών που διενεργούνται μέσω διαδικτύου με αποκλειστικότητα για 5 έτη.

Στόχοι της εταιρείας Ιπποδρομίες Α.Ε. είναι η άρτια διοργάνωση της ιπποδρομιακής δραστηριότητας στο Μαρκόπουλο Αττικής, η ανάπτυξη των ελληνικών ιπποδρομιών, η περαιτέρω διάδοση του αθλήματος στην Ελλάδα, καθώς και η προσφορά στο κοινό μίας σύγχρονης παικτικής εμπειρίας τόσο σε ελληνικές όσο και διεθνείς ιπποδρομίες, σύμφωνα με τις πιο σύγχρονες αρχές του υπεύθυνου παιχνιδιού – για την εφαρμογή τους η ΟΠΑΠ Α.Ε. πιστοποιήθηκε πρόσφατα από τον διεθνή οργανισμό World Lottery Association (WLA).

Διοικητικό Συμβούλιο

Το Διοικητικό Συμβούλιο της εταιρείας Ιπποδρομίες Α.Ε. αποτελείται από τους:

Damian Cope	Πρόεδρος
Σπυρίδων Φωκάς	Α' Αντιπρόεδρος
Οδυσσέας Χριστοφόρου	Β' Αντιπρόεδρος
Kamil Ziegler	Μέλος
Michal Houst	Μέλος
Ian Catchick	Μέλος
Κωνσταντίνος Αλεξόπουλος	Μέλος
Petr Matejovsky	Μέλος

8.2 Περιγραφή της θέσης σας στο τμήμα (ή τμήματα)

Την περίοδο που έκανα την πρακτική μου στις Ιπποδρομίες Α.Ε. (Σεπτέμβριος 2019 – Ιανουάριος 2020) ο οργανισμός βρισκόταν σε φάση αναδιοργάνωσης. Το κύριο μέλημα της εταιρείας στην παρούσα φάση ήταν η ολοκλήρωση της συγχώνευσης με την εταιρεία που την εξαγόρασε, δηλαδή τον ΟΠΑΠ Α.Ε.

Η συγχώνευση έπρεπε να γίνει τόσο σε επίπεδο διαδικασιών όσο και σε τεχνικό. Κατά τη διάρκεια αυτή κάποια τμήματα συγχωνεύθηκαν ή καταργήθηκαν και υπήρχε αντίστοιχη κινητικότητα των εργαζομένων μεταξύ των τμημάτων με νέα καθήκοντα και αρμοδιότητες.

Το Τμήμα απασχόλησης μου ήταν το Racing Department (Ιπποδρομιακό τμήμα). Η κύρια αρμοδιότητα ήταν η καθημερινή βοήθεια εσωτερικών εργασιών, ανάλυση δεδομένων - στατιστικών και η διαχείριση έργου (project management) σε μια από τις καλύτερες εταιρείες της χώρας μας προκειμένου να εφαρμόσω όλα όσα είχα διδαχθεί στο Μεταπτυχιακό ενώ παράλληλα να σχετίζετε με τον χώρο του αθλητισμού λόγω του πρώτου μου πτυχίου στο Sports Management.

Στο τμήμα Racing υπάρχουν εργαζόμενοι - αρμόδιοι για πολλά θέματα όσο αναφορά τις ιπποδρομιακές συγκεντρώσεις. Η θέση μου δεν ήταν μόνο να διαχειριστώ ένα έργο καθαυτό, αλλά η στήριξη των συνάδελφων μου παράλληλα είτε ομαδικά είτε σε ατομικό επίπεδο για τα διάφορα project που έτρεχε το συγκεκριμένο τμήμα την περίοδο της πρακτικής μου.

9 Έργα και Δραστηριότητες

9.1 Έργα, Δραστηριότητες και Προγραμματισμός

Οι συνάδελφοι του τμήματος με ενημέρωσαν για τα καθήκοντά τους και τις καθημερινές πρακτικές που ακολουθούν στην εργασία τους. Μου έκαναν μια γρήγορή ξενάγηση στις εγκαταστάσεις του Ιπποδρόμου. Δύστυχος δεν πέρασα από το πρόγραμμα εισαγωγής των νέων εργαζομένων στον οργανισμό (induction program), γιατί μάλλον αμέλησαν την παρουσία μου. Ενημερώθηκα από άλλους συναδέλφους πως αυτό έχει διάρκεια 1 εβδομάδα (Δευτέρα – Παρασκευή), κατά τη διάρκεια της οποίας οι νέοι εργαζόμενοι παρακολουθούν ενημερωτικές διαλέξεις από στελέχη της εταιρείας, έτσι ώστε να εγκλιματιστούν γρηγορότερα και να κατανοήσουν με σαφήνεια τη δομή της εταιρείας και τις αξίες της.

Ακόμη, για να καταφέρω να καταλάβω καλύτερα τη λειτουργία του τμήματος με ενημέρωσαν για τις εταιρίες που εμπλέκονται στις δημοπρασίες αλόγων (Tattersalls, Goffs και Arqana).

Μου δόθηκαν ακόμα για να μελετήσω οι γενικές διατάξεις της ΦΕΕ, ο κώδικας ιπποδρομιών από τη γενική γραμματεία αθλητισμού, οι διαδικασίες εγγραφών ίππων στις κούρσες, η πολιτική επάθλων, ο οδηγός για το Handicapping και κάποιές άλλες γενικές πληροφορίες που θα πρέπει να γνωρίζω για τον χώρο που εργάζομαι. Τα κύρια έργα που ανέλαβα μετά τον εγκλιματισμό μου στην εταιρεία αναφέρονται αναλυτικά παρακάτω.

9.2 Ανάλυση Κύριων Έργων/Δραστηριοτήτων

Οι βασικές αρμοδιότητες και έργα που ανέλαβα ήταν τα εξής:

- Δημιουργία βάσης δεδομένων στο Microsoft Access για καλύτερη εσωτερική οργάνωσης της αλληλογραφίας του τμήματος.
- Δημιουργία φύλλου στο Excel για τον καθορισμό των συγκεκριμένων αρμοδιοτήτων και υποχρεώσεων κάθε ομάδας μέσα στο τμήμα , αλλά και για κάποιους συνάδελφους μεμονωμένα.
- Συνάντηση (meeting) με τα διοικητικά στελέχη προκειμένου να παραθέσουμε επιγραμματικά τις ανάγκες που έχουμε (ως προς τα αρμόδια άτομα των τμημάτων) προκειμένου να διεξαχθεί με επιτυχία το Ντέρμπι (ιπποδρομιακός αγώνας).
- Ανάλυση δεδομένων ως προς τις Αναφορές Εγγραφών και Αναφορές Αποτελεσμάτων κατά τη διάρκεια του πιο σημαντικού αγώνα της χρονιάς , στο 73ο Ελληνικό Ντέρμπι του ΟΠΑΠ
- Τήρηση πρακτικών των συναντήσεων σε συνεργασία με άλλους εργαζόμενους του τμήματος.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Παρακολούθηση της διαδικασίας εγγραφών των ίππων από τους προπονητές.
- Ενημέρωση σχετικά με τις διαδικασίες που γίνεται διαχωρισμός των ίππων ανάλογα με τα ratings.
- Ενημέρωση σχετικά με το νέο τρόπο διαχωρισμού των ίππων σε κατηγορίες , ο οποίος θα ισχύει για τις κούρσες που θα διεξαχθούν από το 2020 και μετά.
- Ενημέρωση σχετικά με τον νέο τρόπο που θα γίνεται η εγγραφή των ίππων από το 2020 και μετά.
- Ανάλυση και παρουσίαση δεδομένων σχετικά με τους συμμετέχοντες και τα αποτελέσματα των ιπποδρομιών του παρελθόντος σε συναδέλφους. Οι αναλύσεις με βάση το excel γινόντουσαν και με σκοπό να βοηθήσω το τμήμα να οργανώσει καλύτερα τις ιπποδρομιακές συγκεντρώσεις που ορίζουν στα τριμηνιαία προγράμματα των κουρσών ανάλογα τη ζήτηση που υπήρχε στο παρελθόν από τους προπονητές.
- Παροχή βοήθειας σε συναδέλφους για την έγκαιρη ολοκλήρωση projects.
- Εσωτερική ενημέρωση των διοικητικών στελεχών του Ιπποδρόμου για τους ίππους που βρίσκονται σε αποχή και τους λόγους της αποχής.
- Συμβολή στην εσωτερική αναδιοργάνωση των διαδικασιών. Η προσπάθεια του τμήματος ήταν να γίνει νέα κατανομή των αρμοδιοτήτων του τμήματος στους εργαζομένους με σκοπό την απόδοση των εργαζομένων και του τμήματος συνολικά στο 100 %.
- Συλλογή δεδομένων και έλεγχος του Newsletter, στο οποίο περιλαμβάνεται ρεπορτάζ, φωτογραφικό υλικό και αποτελέσματα από τον χώρο του ιπποδρόμου σε εβδομαδιαία βάση.
- Ενημέρωση σχετικά με το καινούριο ιστότοπο που θα αναρτιούνται ηλεκτρονικά τα αποτελέσματα των κουρσών κάθε ιπποδρομιακής συγκέντρωσης.
- Συλλογή δεδομένων για το Factbook του 2019. Στόχος της εταιρείας ήταν η δημοσίευση ενός τεύχους-σύνοψης της χρονιάς που απευθύνεται στη φίλιππη κοινότητα , στα πρότυπα άλλων μεγάλων ιπποδρόμων του εξωτερικού.

9.3 Παραδοτέα του έργου

Πολλά από τα εργαλεία και τα δεδομένα που χρησιμοποιήθηκαν είναι διαθέσιμα για εσωτερική χρήση από τις ομάδες και τους υπαλλήλους της εταιρείας. Στο παράρτημα υπάρχει ένα δείγμα από το εβδομαδιαίο Newsletter της Ιπποδρομίες Α.Ε. καθώς και από ένα πρότυπο Factbook του ιπποδρόμου της Ιρλανδίας το οποίο ακολουθεί η εταιρεία ως πρότυπο. Η προσπάθεια για τη δημιουργία Factbook δεν ευοδώθηκε, όμως άμεσο στόχο της εταιρείας είναι να δημιουργήσεις στο τέλος του τρέχοντος έτους- αρχές του επόμενου το Factbook του 2020.

10 *Γνώση, Δεξιότητες και Κριτική αξιολόγηση*

10.1 Γνώσεις και Δεξιότητες

Γενικότερα η εμπειρία της πρακτικής άσκησης ήταν πολύ θετική στην Ιπποδρομίες Α.Ε. Αυτό συνέβη γιατί ήταν η πρώτη φορά που συμμετείχα ενεργά σε μια εταιρεία που έχει τη δομή, την οργάνωση και τις λειτουργίες μιας εταιρείας Παγκόσμιας εμβέλειας. Η τριβή σε καθημερινή βάση με τους συναδέλφους από τα διάφορα τμήματα της εταιρείας αλλά και με τους προϊστάμενους με βοήθησαν στις κοινωνικές δεξιότητες και στην προσαρμογή σε ομαδικό περιβάλλον, κάτι που θεωρώ πολύ σημαντικό στοιχείο για την έναρξη της επαγγελματικής μου σταδιοδρομίας.

10.2 Κριτική και Προτάσεις

Σχετικά με την νοοτροπία της εταιρείας, θα ήθελα να επισημάνω πως υπάρχουν πολλά σημεία αδιαφάνειας σε εσωτερικές διαδικασίες και διαδικασίες αλλαγής θέσης / προαγωγής ή μετακίνησης. Ακόμα κάποιες μέθοδοι που χρησιμοποιούνται σε μερικά τμήματα γίνεται με παρωχημένες τεχνολογίες, πολλές φορές επειδή η εκπαίδευση του προσωπικού σε πιο σύγχρονες τεχνολογίες δεν κρίνεται απαραίτητη από τη διοίκηση.

Γενικότερα όμως, μετά τη φάση της εξαγοράς από τον ΟΠΑΠ η εταιρεία προσπαθεί να προσαρμοστεί στις αξίες του ΟΠΑΠ Α.Ε, που είναι οι εξής:

- Διασκέδαση
- Δυναμισμός
- Πάθος
- Ακεραιότητα

Γίνεται προσπάθεια από όλη την εταιρεία να πρεσβεύει όλα τα παραπάνω, και με ενέργειες και δράσεις που αναλαμβάνει η ΟΠΑΠ προσπαθεί να επικοινωνήσει το ίδιο και στο κοινωνικό σύνολο.

Έχει δημιουργήσει μεγάλες επενδύσεις τόσο σε γραφεία και υποδομές όπως ο Ιππόδρομος και είναι πάντα ανοιχτή σε προτάσεις που μπορούν να εξασφαλίσουν καλύτερες συνθήκες στους εργαζομένους της. Τόσο σε επίπεδο εξοπλισμού (hardware), όσο και σε εργαλεία και λογισμικά (software) για την ευκολότερη διεκπεραίωση των καθημερινών στόχων τους. Κατά τη διάρκεια της πρακτικής μου είχα την ευκαιρία να εργαστώ μαζί με αυτή την ομάδα σε έναν όμιλο εταιρειών που είναι κάθε χρονιά υποψήφιος για το βραβείο Best Place to Work.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

11 *Επίλογος*

Συνοπτικά , η πρακτική στον ΟΠΑΠ ήταν μια πολύ εποικοδομητική εμπειρία, που σε συνδυασμό με το πρόγραμμα των μαθημάτων του Μεταπτυχιακού έδωσαν μια ιδιαίτερη προστιθέμενη αξία στο πρόγραμμα καθαυτό. Κατά τη διάρκεια της πρακτικής γνώρισα πολλά αξιόλογα άτομα και είχα την ευκαιρία να συνεργαστώ με επαγγελματίες που με βοήθησαν να καταλάβω καλύτερα το καθημερινό αντικείμενο του να δουλεύεις στον ιππόδρομο, αλλά και για την περαίωση της διπλωματικής μου εργασίας.

11.1 Σύνοψη και συμπεράσματα

Η διπλωματική εργασία κλήθηκε να ρίξει φως σε ένα πρόβλημα που απασχολεί τον άνθρωπο από καταβολής κόσμου· την πρόβλεψη γεγονότων. Το πρόβλημα δεν είναι κάτι νέο, όμως οι προκλήσεις ανεβαίνουν ανάλογα με τα μέσα που διαθέτουμε για την επίτευξη ακριβέστερων προβλέψεων. Το μοντέλο πρόβλεψης μπορεί να είναι επιτυχημένο και να συμβάλλει στην σωστή δημιουργία προβλέψεων· δεδομένου ότι υπάρχουν τα κατάλληλα εργαλεία και δεδομένα για την υλοποίηση της ιδέας.

Παρ' όλα αυτά , κανένα μοντέλο δεν είναι 100% επιτυχές, και αυτό γιατί η έκβαση της ιπποδρομίας είναι ένα βαθιά στοχαστικό πρόβλημα. Όσο σωστές και να είναι οι τεχνικές μας· και όσο απεριόριστα και λεπτομερή να είναι τα δεδομένα μας ο παράγοντας τύχη είναι πάντα αστάθμητος σε αθλητικά γεγονότα. Η ανθρώπινη φύση πάντα θα προσπαθεί να ξεπεράσει τον εαυτό της και να βρει τρόπους να φτιάχνει ακόμα καλύτερα και πιο λεπτομερή μοντέλα , ιδιαίτερα όσο τα αθλητικά γεγονότα βρίσκονται στο προσκήνιο για τη διασκέδαση μεγάλης μάζας του πληθυσμού σε παγκόσμια κλίμακα.

11.2 Μελλοντικές επεκτάσεις

Η διπλωματική θα μπορούσε να εμπλουτιστεί στο μέλλον, εάν υπήρχε η πρόσβαση σε ακόμα περισσότερα δεδομένα. Υπάρχουν πολλές μεταβλητές που δεν έχουν ληφθεί υπόψιν για τη δημιουργία του μοντέλου, ενώ ακόμα και το παρών μοντέλο δεν έχει δοκιμαστεί σε έναν ικανό αριθμό αγώνων. Η εφαρμογή του θα μπορούσε με κάποιες τροποποιήσεις να γίνει σε αγώνες του εξωτερικού (και όχι μόνο στην Ελλάδα) ενώ θα μπορούσε να συνδυαστεί και με ένα κατάλληλο frontend έτσι ώστε να είναι περισσότερο user friendly και να είναι πιο εύκολα στη διαχείριση τόσο τα δεδομένα εισόδου, όσο και τα αποτελέσματα. Ακόμα θα μπορούσαν κάποιες εσωτερικές λειτουργίες της εταιρείας να βελτιωθούν και το τμήμα των Αγώνων να δουλεύει πολύ καλύτερα στο μέλλον και να διαφημίσει σε μεγαλύτερο κοινό ένα παρεξηγημένο άθλημα όπως είναι οι Ιπποδρομίες. Κλείνοντας, μπορούν να γίνουν στατιστικά χώρου· διαχωρισμός του σταδίου σε X και Y άξονα και ουσιαστικά να δημιουργούνται συντεταγμένες. Με βάση αυτές τις συντεταγμένες, να καταγράφεται το σημείο του τερματισμού και μετά αναλύεται, συνήθως με τη χρήση R.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

12 *Βιβλιογραφία*

[McF74] McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In: P. Zarembka (Ed.), Frontiers in Econometrics. New York: Academic Press, pp.105-142.

[JJT06] Johnson, J. E. V., Jones, O. & Tang, L. (2006). Exploring decision makers' use of price information in a speculative market. Management Science, 52(6), 897-908

[Ede07] Edelman, D. (2007). Adapting support vector machine methods for horserace odds prediction. Annals of Operations Research, 151(1), 325-336.

[LSJ08] Lessmann, S., Sung, M. & Johnson, J. E. V. (2008). Identifying winners of competitive events: An SVM-based classification model for horserace prediction. European Journal of Operational Research, (doi:10.1016/j.ejor.2008.03.018).

[CS82] Chapman, R. G. & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. Journal of Marketing Research, 19(3), 288-301

[BC86] Bolton, R. N. & Chapman, R. G. (1986). Searching for positive returns at the track: A multinomial logit model for handicapping horseraces. Management Science, 32(8), 1040-1060.

[GHB03] Gu M. G., Huang, C. & Benter, W. (2003). Multinomial probit models for competitive horse racing. Working paper, Chinese University of Hong Kong.

[Vap95] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. New York: Springer. Zar, J. H. (2007). Biostatistical Analysis. Upper Saddle River: Prentice Hall.

[CN06] Caruana, R. & Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms. In: W. W. Cohen & A. Moore (Eds.), Proc. of the 23rd Intern. Conf. on Machine Learning. New York: ACM.

[CLB04] Chen, C., Liaw, A. & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. Technical Report, 666, Statistics Department, University of California at Berkeley.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

[HL05] Hamza, M. & Larocque, D. (2005). An empirical comparison of ensemble methods based on classification trees. Journal of Statistical Computation and Simulation, 75(8), 629-643.

[Bre01] Breiman, L. (2001). Random forests. Machine Learning, 45(1).

[BFO84] Breiman, L, Friedman, J. H., Olshen, R. & Stone, C. (1984). Classification and Regression Trees, Belmont: Wadsworth

[SJ07] Sung, M. & Johnson, J. E. V. (2007). Comparing the effectiveness of one- and two-step conditional logit models for predicting outcomes in a speculative market. Journal of Prediction Markets, 1(1), 43-59

13 *Παράρτημα*

13.1 Κώδικας r

```
##φέρνουμε κάποια libraries από το repository
library(plyr)
library(dplyr)
library(stringr)
library(caret)

#πρώτα βάζω τα dr files
setwd("")

#data από άλογα από arxeio .csv
temp = list.files(pattern = "")
myfiles = lapply(temp, read.csv, header = FALSE)
myfiles2 = lapply(temp, read.csv, header = FALSE, quote = "")
HorseData1 = ldply(myfiles, data.frame)
program = ldply(myfiles2, data.frame)
HorseData1 =
HorseData1[,c(1,2,3,4,19,21,24,25,41,43,64,65,69,71,75,189,201,202)]
names(HorseData1) = c("Track", "RDate", "RNum", "PP",
"TrainRecordCurYear", "JockRecordCurYear", "HName", "YOB",
"Medication",
                      "EquipChange", "Year", "StartsYear", "EarnYear",
"StartsPrev", "EarnPrev", "BRISRunstyle", "ProgramNum", "MLOdds")
x = str_replace(program[,201], "[:punct:]", "")
HorseData1$ProgramNum = str_replace(x, "[:punct:]", "")

#παλία αποτελέσματα από αρχείο .csv
temp = list.files(pattern = "")
myfiles = lapply(temp, read.csv, header = FALSE)
PastPerf1 = ldply(myfiles, data.frame)
PastPerf1 = PastPerf2[,c(1,2,3,4,5,10,11,14,35,41,53,56,57,61,64,84)]
```

```
names(PastPerf1) = c("Horse Name","Age", "Sex",
"Equip","Trainer","Jockey", "Hand.","W.-IN","Time","Margins","Stall")

#αρχεία BRIS from .csv
temp = list.files(pattern = "")
myfiles = lapply(temp, read.csv, header = FALSE)
BRISStats = ldply(myfiles, data.frame)
BRISStats = BRISStats[,c(1,2,3,4,5,10)]
names(BRISStats) = c("Track", "RDate", "RNum")

ls()

#Race data
RaceData$PostTime = gsub("[\\(, \\), \\/]", "", substr(RaceData$PostTime,
1, 7))

#Horse data

##ποσοστό νίκης προπονητή
HorseData$TrainStartCurYear =
as.numeric(substr(HorseData$TrainRecordCurYear, 1, 4))
HorseData$TrainWinCurYear =
as.numeric(substr(HorseData$TrainRecordCurYear, 5, 8))
HorseData$TrainWinPct =
HorseData$TrainWinCurYear/(HorseData$TrainStartCurYear + 1) # +1 gia
na mhn exo undefined

##ποσοστό νίκης  jockey
HorseData$JocStartCurYear =
as.numeric(substr(HorseData$JockRecordCurYear, 1, 4))
HorseData$JocWinCurYear =
as.numeric(substr(HorseData$JockRecordCurYear, 5, 8))
HorseData$JockWinPct =
HorseData$JocWinCurYear/(HorseData$JocStartCurYear + 1)


##Equipment changes
HorseData$addBlinks = ifelse(HorseData$EquipChange == 1, 1, 0)
HorseData$remBlinks = ifelse(HorseData$EquipChange == 2, 1, 0)

HorseData = select(HorseData, Track, RDate, RNum, PP, HName,
BRISRunstyle, TrainWinPct, JockWinPct, logEarn, addBlinks, remBlinks,
MLOdds, ProgramNum)

##Δεδομένα παλιών επιδόσεων

##η θέση τερματισμού περιλαμβάνει άλογα που δεν τελείωσαν τη κούρσα
##σε αυτή την περίπτωση λέμε πως τερμάτισαν τελευταία
##εάν δεν υπάρχει χρόνος θεωρώ ότι δεν τελείωσαν τη, δεν με ενδιαφέρει ο
λόγος
PastPerf$FinishPos =
ifelse(is.na(as.numeric(as.character(PastPerf$FinishPos))),
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
                    PastPerf$NumHor,
                    as.numeric(as.character(PastPerf$FinishPos)))
PastPerf$FinishPos = ifelse(PastPerf$FinishPos %in% c(0,89,92,99),
PastPerf$NumHor, PastPerf$FinishPos)

##το πεδίο Margin είναι είτε απόσταση από το προπορευόμενο ίππο είτε
περιθώριο νίκης, ανάλογα εάν το άλογο κέρδισε ή όχι
PastPerf$BL = as.numeric(ifelse(PastPerf$FinishPos == 1, 0,
PastPerf$BLMargin))
PastPerf$Margin = as.numeric(ifelse(PastPerf$FinishPos == 1,
PastPerf$BLMargin, 0))

##πρώτη ενδιαφέρουσα μεταβλητή, αξιολογώ τους ίππους από ποσοστό
τερματισμού, τα πλασέ, και τις νίκες/ήττες τους
##θα μπορούσα να βάλλω και άλλες μεταβλητές, όμως το κάνω πιο απλό

PastPerf$raceRating = sqrt((1 -
PastPerf$FinishPos/PastPerf$NumHor)*(log(PastPerf$Purse)*(log(1+PastP
erf$Margin)+log(1+PastPerf$BL))))

##ο κάθε ίππος βαθμολογείτε σε 365 ημέρες για τη φόρμα του
##η κατηγοριοποίηση του ίππου μπορεί να γίνει με δεδομένα μέχρι 365
ημέρες πίσω, αυτό το χρονικό περιθώριο δίνω γιατί μέχρι εκεί έχω
δεδομένα.

## βλέπω αποτελέσματα κούρσας  στις 120 πιο πρόσφατες ημέρες
PastPerf[is.na(PastPerf)] = 0
pastPerf120 = filter(PastPerf, DaysOld <= 120)
pastPerf120 = group_by(pastPerf120, Track, RDate, RNum, PP)
pastPerf120 = summarise(pastPerf120, countRaces120 = length(Track),
avgBL = mean(BL), avgRaceRating = mean(raceRating))

##βλέπω αποτελέσματα κούρσας στις  365 πιο πρόσφατες ημέρες
pastPerf365 = filter(PastPerf, DaysOld <= 365, !is.na(BRISSpeed))
pastPerf365 = group_by(pastPerf365, Track, RDate, RNum, PP)
pastPerf365 = summarise(pastPerf365, avgSpeed = mean(BRISSpeed),
avg2FSpeed = mean(BRIS2fPace),
                    avg4FSpeed = mean(BRIS4fPace), avgLateSpeed =
mean(BRISLatePace),
                    turfRating = (1-(length(PrevRDate) - sum(PrevCondition
%in% c('FM', 'YL')))/length(PrevRDate))*
                      sum(PrevCondition %in% c('FM', 'YL') & FinishPos
<=3)/sum(PrevCondition %in% c('FM', 'YL')),
                    avgRaceRating365 = mean(raceRating))


##συνδυάζω τα δεδομένα
PastPerf = left_join(PastPerf365, pastPerf120)
PastPerf[is.na(PastPerf)] = 0
PastPerf$countsq = PastPerf$countRaces120*PastPerf$countRaces120
## τα βάζω όλα μαζί
h = inner_join(HorseData1, BRISStats)
```

```
h = inner_join(h, RaceData)
h = left_join(h, PastPerf)
h = unique(h)
h[is.na(h)] = 0
h$Track = str_trim(h$Track)

## υπολογίζω τον ανταγωνισμό και τον ίππο μέσα στον αγώνα
i = group_by(h, Track, RDate, RNum)
j = summarise(i, meanEarn = mean(logEarn), meanPurse = mean(logPurse),
          meanSpeed = mean(avgSpeed), meanES = mean(avg2FSpeed),
mean4f = mean(avg4FSpeed),  meanLS = mean(avgLateSpeed),
          numHor = length(PP), meanPower = mean(PowerRating),
meanTurfRat = mean(turfRating), maxLevel = max(highestLevel),
          meanRaceRat365 = mean(avgRaceRating365), meanRaceRat =
mean(avgRaceRating), meanBL = mean(avgBL))
h = inner_join(i, j)

h$compTurfRat = ifelse(h$meanTurfRat == 0, 0, h$avgTurfRating -
h$meanTurfRat)
h$compRaceRat = ifelse(h$meanRaceRat == 0, 0, h$avgRaceRating -
h$meanRaceRat)
h$compRaceRat365 = ifelse(h$meanRaceRat365 == 0, 0,
h$avgRaceRating365 - h$meanRaceRat365)
h$compLevel = ifelse(h$maxLevel == 0, 0, h$highestLevel - h$maxLevel)
h$compBL = ifelse(h$meanBL == 0, 0, h$avgBL - h$meanBL)
h$compEarn = ifelse(h$logEarn == 0, 0, h$logEarn - h$meanEarn)
h$compPurse = ifelse(h$logPurse == 0, 0, h$logPurse- h$meanPurse)
h$compSpeed = ifelse(h$avgSpeed == 0, 0, h$avgSpeed - h$meanSpeed)
h$compES = ifelse(h$avg2FSpeed == 0, 0, h$avg2FSpeed - h$meanES)
h$comp4f = ifelse(h$avg4FSpeed == 0, 0, h$avg4FSpeed - h$mean4f)
h$compLS = ifelse(h$avgLateSpeed == 0, 0, h$avgLateSpeed - h$meanLS)
h$compPower = ifelse(h$PowerRating == 0, 0, h$PowerRating -
h$meanPower)
h$raceID = factor(paste(h$Track, h$RDate, h$RNum, sep = ""))

##καθορίζω το  workspace

rm(BRISStats, PastPerf, HorseData1, RaceData, i, j,  pastPerf365,
pastPerf120, speedPC, classPC)

ls()

##πρέπει να συμπεριλάβω τα δεδομένα στον αλγόριθμο, μετά θα χωρίσω το
trainset και το testset
h$Track = str_trim(as.character(h$Track))
Results$Track = str_trim(as.character(Results$Track))
h$RDate = as.character(h$RDate)
Results$RDate = as.character(Results$RDate)
h$HName = toupper(h$HName)
Results$HName = toupper(Results$HName)
Results$HName = str_trim(gsub("\\(.*$", "", Results$HName ))
h = inner_join(h,Results, by = c("Track", "RNum", "RDate", "HName"))
```

```
rm(Results)

ls()
##########
##ορίζω το trainset ως ποσοστό του συνολλικού data set – γίνετε σε επίπεδο
race και όχι horse
trainRaces = sample(unique(ds$raceID), size =
floor(length(unique(ds$raceID))*.1))

dsTrain = filter(ds, raceID %in% trainRaces)
dsTest = filter(ds, !(raceID %in% trainRaces))
# όσο μικρότερο είναι το μήκος  που νικάει ένα άλογο, τόσο καλύτερο
είναι. Χρησιμοποιώ λογαριθμική κλίμακα όμως για να μπορέσω να δώσω
μεγαλύτερη αξία στις μικρές διαφορές.
#μεταξύ των ίππων – δεν είναι το ίδιο να έχει διαφορά 1 με 2 από τη
διαφορά 8 με 9.

dsBLModel = train(log(finishBL+1) ~ compSpeed + compES + comp4f +
compLS + BestSpeedLife + compEarn + compPurse + compLevel +
                  compRaceRat + TrainWinPct + JockWinPct + numHor +
compPower + BRISRunstyle + countRaces120 + countsq,
                method = 'gbm', data = dsTrain,verbose = F)

dsBLModel

dsTrain$preBL = predict(dsBLModel, dsTrain)

#κάνω προσομοίωση Monte Carlo
monteCarlo = function(val, StD, n) {
  result = val + rnorm(n, mean = val, sd = StD)
  return(result)
}

simDayMin = function(x, StD, n) {
  results = vector("numeric", 0)
  for (i in 1:length(unique(x$raceID))) {
    xrace = subset(x, x$raceID == as.character(unique(x$raceID))[i])
    sim = vector()
    for (j in 1:length(xrace$PP)) {
      sim = cbind(sim, monteCarlo(as.numeric(xrace[j,3]), StD, n))
    }
    winners = as.vector(1:length(xrace$raceID))
    for (k in 1:n) {
      winners = c(winners, which.min(sim[k,]))
    }
    results = c(results, as.vector(table(winners)))
  }
  return(results)
}
```

```
#Sim results with different variances
dsTrain = dsTrain[order(dsTrain$raceID, dsTrain$PP.y),]
simTrain = dsTrain[,c("raceID", "PP.y", "preBL")]
dsTrain$wins871 = simDayMin(simTrain[,c(1,2,3)], .871, 20000) -1
dsTrain$wins721 = simDayMin(simTrain[,c(1,2,3)], .721, 20000) -1
dsTrain$wins1021 = simDayMin(simTrain[,c(1,2,3)], 1.021, 20000) -1
dsTrain$wins1171 = simDayMin(simTrain[,c(1,2,3)], 1.171, 20000) -1
dsTrain$wins1321 = simDayMin(simTrain[,c(1,2,3)], 1.321, 20000) -1
dsTrain$wins1471 = simDayMin(simTrain[,c(1,2,3)], 1.471, 20000) -1
dsTrain$wins1621 = simDayMin(simTrain[,c(1,2,3)], 1.621, 20000) -1
dsTrain$wins1771 = simDayMin(simTrain[,c(1,2,3)], 1.771, 20000) -1


##Κατηγοριοποιώ τους ίππους ανάλογα με το ποσοστό νίκης τους
##χρησιμοποιώ 4 group με περίπου ίδιο αριθμό ίππων στο καθένα
library(Hmisc)
dsTrain$group721 = cut2(dsTrain$wins721, g = 4)
dsTrain$group871 = cut2(dsTrain$wins871, g = 4)
dsTrain$group1021 = cut2(dsTrain$wins1021, g = 4)
dsTrain$group1171 = cut2(dsTrain$wins1171, g = 4)
dsTrain$group1321 = cut2(dsTrain$wins1321, g = 4)
dsTrain$group1471 = cut2(dsTrain$wins1471, g = 4)
dsTrain$group1621 = cut2(dsTrain$wins1621, g = 4)
dsTrain$group1771 = cut2(dsTrain$wins1771, g = 4)
################################################################
########################################
################################################################
########################################
##γράφημα για το πόσο καλά έγινε η καθαριοποίηση
x = group_by(dsTrain, group1771)
x = summarise(x, trueWinPct = sum(FinishPos == 1)/length(raceID),
predWinPct = mean(wins1771)/20000)

qplot(x$predWinPct, x$trueWinPct) + geom_smooth(method = 'lm')
################################################################
########################################
##υπολογισμός κερδοφορίας – αξιολόγηση testset
## εισαγωγή του αρχείου με τις αποδόσεις σε se .csv
temp = list.files(pattern = ".odds")
myfiles = lapply(temp, read.csv, header = FALSE)
myfiles2 = lapply(temp, read.csv, header = FALSE, quote = "")
OddsData1= ldply(myfiles, data.frame)
dsTest = dsTest[order(dsTest$raceID, dsTest$PP.y),]
simTest = dsTest[,c("raceID", "PP.y", "preBL")]
dsTest$wins1321 = simDayMin(simTest[,c(1,2,3)], 1.321, 20000) -1

dsTrain$fairOdds = 20000/dsTrain$wins1321 - 1
dsTest$fairOdds = 20000/dsTest$wins1321 - 1

dsTrain$betSize = ((dsTrain$wins1321/20000)*(dsTrain$FOdds +1)-
1)/dsTrain$FOdds
```

```
dsTest$betSize = ((dsTest$wins1321/20000)*(dsTest$FOdds +1)-
1)/dsTest$FOdds


#βασικό στοίχημα, εάν οι αποδόσεις είναι μεγαλύτερες από αυτές που
υπολογίζω τοποθετώ στοίχημα
dsTrain$bet1 = ifelse(dsTrain$fairOdds <= dsTrain$FOdds,
ifelse(dsTrain$FinishPos == 1,dsTrain$FOdds*dsTrain$betSize,-
dsTrain$betSize), 0)

##
##εάν η αγορά αξιολογεί ένα άλογο 50 % κάτω από την πραγματική του
τιμή, θα το ποντάρω εάν δίνει κάτω από 15.00 απόδοση
##η μεταβαλότητα ανεβαίνει σε μεγαλύτερες αποδόσεις


#το ίδιο και για το testset
dsTest$bet1 = ifelse(dsTest$fairOdds <= dsTest$FOdds,
ifelse(dsTest$FinishPos == 1,dsTest$FOdds*dsTest$betSize,-
dsTest$betSize), 0)

dsTest$bet2 = ifelse(dsTest$fairOdds <= 1.5 & dsTest$fairOdds <=
dsTest$FOdds,
                ifelse(dsTest$FinishPos == 1,dsTest$FOdds*dsTest$betSize,-
dsTest$betSize), 0)

dsTrain$payout = dsTrain$bet1+dsTrain$bet2
dsTest$payout = dsTest$bet1+dsTest$bet2


roiTrain = 1 + sum(dsTrain$payout)/sum(abs(dsTrain$betSize))

roiTest = 1 + sum(dsTest$payout)/sum(abs(dsTest$betSize))

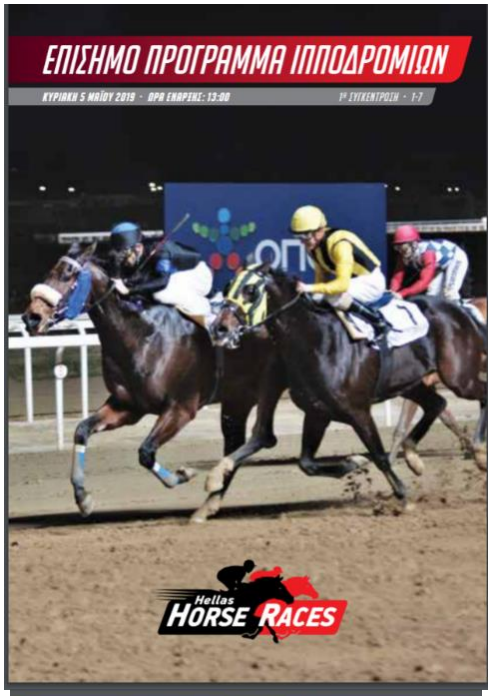

roiTest
## το αποτέλεσμα σε μέσο όρο είναι η ροή = 1.011316
## έχω κέρδος κατά μέσο όρο 1,1
```

13.2 Πηγές δεδομένων [πηγή:

https://www.markopoulopark.gr/]

13.2.1 Επίσημο Πρόγραμμα Ιπποδρομιών

[πηγή: https://www.markopoulopark.gr/schedule]





13.2.2 Αποτελέσματα (ιπποδρομιακών συγκεντρώσεων ανά κούρσα)

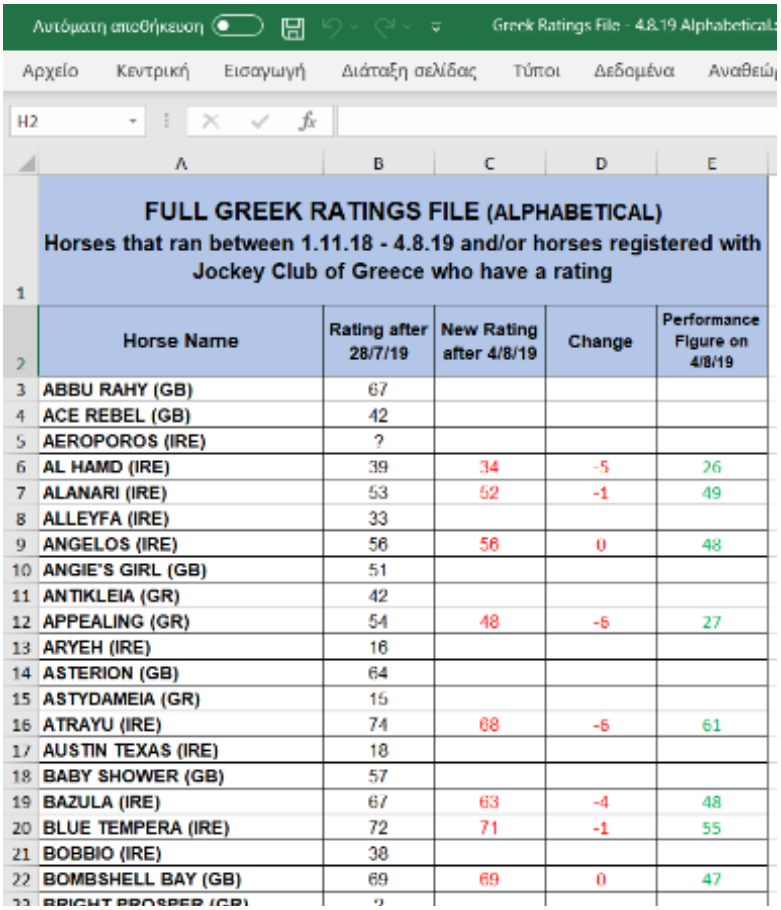
[πηγή: https://www.markopoulopark.gr/results]





13.2.3 Rating ίππων

[πηγή: https://www.markopoulopark.gr/table-of-weights]



	A	B	C	D	E
1	**FULL GREEK RATINGS FILE (ALPHABETICAL)** **Horses that ran between 1.11.18 - 4.8.19 and/or horses registered with Jockey Club of Greece who have a rating**				
2	**Horse Name**	**Rating after 28/7/19**	**New Rating after 4/8/19**	**Change**	**Performance Figure on 4/8/19**
3	ABBU RAHY (GB)	67			
4	ACE REBEL (GB)	42			
5	AEROPOROS (IRE)	?			
6	AL HAMD (IRE)	39	34	-5	26
7	ALANARI (IRE)	53	52	-1	49
8	ALLEYFA (IRE)	33			
9	ANGELOS (IRE)	56	56	0	48
10	ANGIE'S GIRL (GB)	51			
11	ANTIKLEIA (GR)	42			
12	APPEALING (GR)	54	48	-6	27
13	ARYEH (IRE)	16			
14	ASTERION (GB)	64			
15	ASTYDAMEIA (GR)	15			
16	ATRAYU (IRE)	74	68	-6	61
17	AUSTIN TEXAS (IRE)	18			
18	BABY SHOWER (GB)	57			
19	BAZULA (IRE)	67	63	-4	48
20	BLUE TEMPERA (IRE)	72	71	-1	55
21	BOBBIO (IRE)	38			
22	BOMBSHELL BAY (GB)	69	69	0	47
23	BRIGHT PROSPER (GB)	?			

FULL GREEK RATINGS FILE (ALPHABETICAL) **Horses that ran between 1.3.19 - 29.12.19 and/or horses registered with Jockey Club of Greece**					
Horse Name	**Όνομα Ίππου**	**New Rating after 29/12/19**	**Rating after 22/12/19**	**Change**	**Performance Figure on 29/12/19**
ANGELOS (IRE)	ΑΓΓΕΛΟΣ (IRE)	48	48		
AEROPOROS (IRE)	ΑΕΡΟΠΟΡΟΣ (IRE)	29	29		
AIGLI (IRE)	ΑΙΓΛΗ (IRE)	21	21		
I'M WALKING HERE (IRE)	ΑΪΜ ΓΟΥΟΚΙΝΓΚ ΧΙΕΡ (IRE)	47	48	-1	0
IM TALKING TO YOU (IRE)	ΑΪΜ ΤΟΚΙΝΓΚ ΤΟΥ ΓΙΟΥ (IRE)	43	?	first rating	43
IRISH BREEZE (GB)	ΑΪΡΙΣ ΜΠΡΗΖ (GB)	70	70		
AL HAMD (IRE)	ΑΛ ΧΑΜΝΤ (IRE)	43	43		
ALANARI (IRE)	ΑΛΑΝΑΡΙ (IRE)	56	56		
ALLEYFA (IRE)	ΑΛΛΕΪΦΑ (IRE)	30	30		
ALSAFA (GB)	ΑΛΣΑΦΑ (GB)	44	44		
AMOR DE VEGA GB)	ΑΜΟΡ ΝΤΕ ΒΕΓΚΑ (GB)	?	?		
ABBU RAHY (GB)	ΑΜΠΟΥ ΡΑΧΙ (GB)	67	65	2	67
AMBYFAEIRVINE (IRE)	ΑΜΠΥΦΕΡΒΑΪΝ (IRE)	58	58		
ANIL (GR)	ΑΝΙΛ (GR)	unraced	unraced		
ANCIENT CITY (FR)	ΑΝΣΙΕΝΤ ΣΙΤΙ (FR)	65(FR)	65(FR)		
ANGIE'S GIRL (GB)	ΑΝΤΖΥΣ ΓΚΕΡΛ (GB)	54	46	8	63
ANTIKLEIA (GR)	ΑΝΤΙΚΛΕΙΑ (GR)	42	42	0	15

13.2.3.1 Πώς σκέφτεται ο Ισοζυγιστής

[πηγή: https://www.markopoulopark.gr/documents/20143/0/%CE%A0%CF%8E%CF%82+%CE%A3%CE%BA%CE%AD%CF%86%CF%84%CE%B5%CF%84%CE%B1%CE%B9+%CE%BF+%CE%99%CF%83%CE%BF%CE%B6%CF%85%CE%B3%CE%B9%CF%83%CF%84%CE%AE%CF%82%3B+%CE%9F+PHIL+SMITH+%CE%B1%CE%BD%CE%B1%CE%BB%CF%8D%CE%B5%CE%B9+%CF%84%CE%B7%CE%BD+25%CE%B7+%CE%91%CF%85%CE%B3%CE%BF%CF%8D%CF%83%CF%84%CE%BF%CF%85.pdf/578af565-5c7c-acda-56de-1e10349d1bf0]

ΕΠΕΞΗΓΗΣΗ ΑΠΟΦΑΣΕΩΝ ΣΧΕΤΙΚΑ ΜΕ ΤΑ HANDICAP RATINGS ΤΗΣ ΚΥΡΙΑΚΗΣ, 25 ΑΥΓΟΥΣΤΟΥ 2019

Πώς σκέφτεται ο Ισοζυγιστής

1Η ΙΠΠΟΔΡΟΜΙΑ / #97

ΑΦ(ΣΤ)	ΧΡΩΜΑΤΑ	# ΟΝΟΜΑ ΙΠΠΟΥ	ΗΛΙΚΙΑ	ΓΕΝΟΣ	ΕΞΟΠ.	ΠΡΟΠΟΝΗΤΗΣ	ΑΝΑΒΑΤΗΣ	ΧΑΝΤ.	ΜΕΤΑΖ.	ΧΡΟΝΟΣ	ΜΗΚΗ	ΑΠΟΔΟΣΗ
1(5)		2. ΓΚΛΟΜΠΑΛ ΣΙΦΤ (IRE)	3	Εκτ.	Π	Θεολόγου Π.	Χριστοφής Μ.	58	58	01:00:84		7,73
2(3)		1. ΑΕΡΟΠΟΡΟΣ (IRE)	3	Αρσ.	Π	Κρητικός Α.	Καποδίστριας Κ.	58	58	01:00:97	3/4	3,06
3(2)		3. ΣΕΡ ΓΚΑΝΙΑΝ (GR)	3	Εκτ.	Π	Καμούτσης Γ.	Τσέλα Ε.	58	58	01:01:31	2	15,51
4(4)		5. ΓΟΥΟΡΙΛΥ (GB)	2	Αρσ.		Σταύρου Χρ.	Δημητσάνης Π.	54	54	01:01:71	2 1/4	1,73
5(1)		4. ΦΕΙΔΙΑΣ (GR)	3	Εκτ.	Π	Καμούτσης Γ.	Περαντινός Κ.	58	58	01:02:88	7	63,42

Ο **ΣΕΡ ΓΚΑΝΙΑΝ** με rating 22 ήταν ο ίππος «οδηγός» και συνεπώς το Επίσημο Rating του, ήταν και η βαθμολόγηση της επίδοσής του στην συγκεκριμένη κούρσα.

Ο **ΓΚΛΟΜΠΑΛ ΣΙΦΤ** κέρδισε τον ΣΕΡ ΓΚΑΝΙΑΝ διαφορά 2.75 μηκών (2.75*2.5=7 μονάδες), αλλά επειδή κέρδισε με άνεση, θεωρούμε ότι η διαφορά είναι στις 9 μονάδες. Συνεπώς η επίδοσή του είναι 22+9 =31 και αυτό θα είναι και το πρώτο Επίσημο Rating του.

Ο **ΓΟΥΟΡΙΛΥ** ως 2ετής θα έπρεπε να κερδίζει από όλους του 3ετείς, συμπεριλαμβανομένου του ΣΕΡ ΓΚΑΝΙΑΝ, 10 κιλά βάση της Κλίμακας Βαρών Ανά Ηλικία και Απόσταση. Ωστόσο, λόγω των όρων της ιπποδρομίας, οι 3ετείς του χάριζαν μόνο 4 κιλά. Συνεπώς, για να μετρήσουμε την επίδοσή του προσθέτουμε, στην επίδοση του ΣΕΡ ΓΚΑΝΙΑΝ (22) την διαφορά των υπολειπόμενων κιλών που θα έπρεπε να του χαρίζουν (12 μονάδες= 6 κιλά) και αφαιρούμε τη διαφορά των μηκών τους (2.25μήκη*2.5= 6). Στη συγκεκριμένη περίπτωση χρησιμοποιώ 5 μονάδες και όχι 6 για την διαφορά των 2.25 μηκών, καθώς ο ΓΟΥΟΡΙΛΥ μαλάκωσε λίγο προς το τέλος τις προσπάθειές του. Άρα, η επίδοση του ΓΟΥΟΡΙΛΥ είναι 22+12-5=29

13.2.4 Εβδομαδιαία νέα ιπποδρομιών

[πηγή: https://www.markopoulopark.gr/weekly-news]

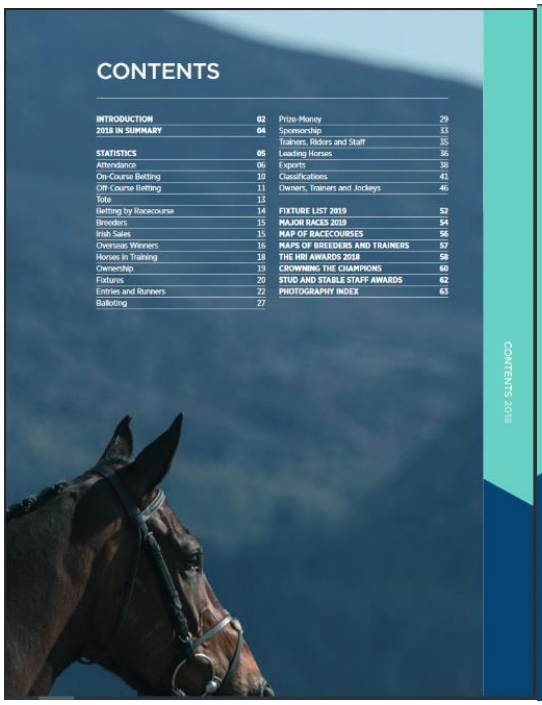




13.2.5 Factbook 2018

[πηγή: https://www.hri.ie/hri/2019/Factbook%202018%20FINAL.pdf]





CONTENTS

INTRODUCTION	02	Prize-Money	29
2018 IN SUMMARY	04	Sponsorship	33
		Trainers, Riders and Staff	35
STATISTICS	05	Leading Horses	36
Attendance	06	Exports	38
On-Course Betting	10	Classifications	41
Off-Course Betting	11	Owners, Trainers and Jockeys	46
Tote	13		
Betting by Racecourse	14	**FIXTURE LIST 2019**	52
Breeders	15	**MAJOR RACES 2019**	54
Irish Sales	15	**MAP OF RACECOURSES**	56
Overseas Winners	16	**MAPS OF BREEDERS AND TRAINERS**	57
Horses in Training	18	**THE HRI AWARDS 2018**	58
Ownership	19	**CROWNING THE CHAMPIONS**	60
Fixtures	20	**STUD AND STABLE STAFF AWARDS**	62
Entries and Runners	22	**PHOTOGRAPHY INDEX**	63
Balloting	27		

INTRODUCTION 2018



On behalf of Horse Racing Ireland, I am pleased to present the fourteenth edition of the Irish Racing Factbook for the year just gone by. These pages contain a wide range of the latest statistics relating to the racing and breeding industries in Ireland, and while the industry continues to perform well, we acknowledge that 2018 was a challenging year.

On the track, at home and abroad, our horses, trainers and jockeys enjoyed tremendous success. There were 17 Irish-trained winners at the Cheltenham Festival, Irish runners filled the first four places in a truly memorable renewal of the Aintree Grand National where Tiger Roll narrowly outpointed Pleasant Company for Gigginstown House Stud, Gordon Elliott and Davy Russell. On the Flat, Aidan O'Brien won three British classics, the Jessica Harrington-trained Alpha Centauri was an outstanding performer, Ken Condon and Joseph O'Brien won their first Classics and Curragh trainer Patrick Prendergast made the Group 1 breakthrough.

Horse Racing Ireland is firmly committed to increasing the numbers involved in racehorse ownership in Ireland. To that end, we are delighted to report an increase of 16.2% year-on-year with almost 800 new owners registering in 2018, and syndicate ownership rising by 33%. The total number of owners in Ireland is up 3.5% and this is also reflected in an increase of 7.4% in the number of horse-in-training. Total entries, across Flat and National Hunt, are up over 3% and while the total number of National Hunt runners has shown a decrease, the number of runners on the Flat is up as are the number of individual winners, which has risen to 1,904 from 1,871 overall.

Prize-money is another priority for us and it reached a figure of €63.5m in 2018, up 3.9% on the 2017 figure of €61.1m. Further increases are budgeted for 2019, totalling €2.9m, with an emphasis on 20 Grade 2 and Grade 3 racecourses around the country which will share €500,000 to enhance their feature races, races that are typically accessible to the majority of the horse population.

It is encouraging that commercial sponsorship in 2018 rose once again. A total input of €5.2m towards the overall prize fund reflects an increase of 5.3%, a significant jump on the 3% rise of the year before. Commercial sponsorship has risen in each of the last three years, confirming horse racing's continued appeal as an advertising and promotional tool for business. The Irish European Breeders Fund sponsorship is up 11.1% to €2.2m and we acknowledge the Fund as the leading sponsor.

Overall, attendances were marginally down to 1.27m in 2018, a drop of 0.5% on the year before. Attendances did show an increase in the latter half of 2018 and we expect that trend to continue, as we prepare for the opening of the new-look Curragh Racecourse which has operated with a reduced capacity for the past two years and we hope to build on the record-breaking crowds seen at the Punchestown Festival in the last two years.



Uncertainty around Brexit was a significant factor as Irish bloodstock sales suffered their first reversal in nine years. A strong market for National Hunt horses was offset by a decline for those selling Flat horses and overall, sales at public auction were down 8% to €161.5m. The value of Irish-foaled exports sold at public auction declined by 1.9% to €263.1, but the number of countries Irish-foaled horses were sold to rose to 33.

After seven years of turnover growth, Tote Ireland suffered a reversal with revenue dropping 31.3% to €19.2m due to a significant falloff in money going into off-course Irish pools from overseas. Total on-course betting saw a combined fall of 9.8% to €75.5m, continuing the downward trend in on-course betting and once again demonstrating the changing nature of the betting markets.

We are facing our challenges head-on. Horse Racing Ireland has been heavily engaged with our European and British counterparts on the matter of Brexit. The racing and breeding industries face ongoing challenges around funding, infrastructure and staffing. Our annual return to the Irish economy tops €1.84bn, while there are almost 29,000 jobs, directly and indirectly, supported by both industries.

If you believe that there are any other areas that should be covered as we aim to produce a comprehensive overview of all the relevant statistics, please let us know.

Brian Kavanagh
Chief Executive

2018 IN SUMMARY



ATTENDANCE 2018

TOTAL ATTENDANCE 2014 - 2018





TOP SEVEN FESTIVALS BY ATTENDANCE 2017 - 2018

Festival	No. of Days 2018	2017	2018	Variance	% +/-
Galway	7	146,332	141,341	-4,991	-3%
Punchestown	5	122,557	127,589	13,032	11%
Listowel	7	89,594	91,393	1,799	2%
Leopardstown (Christmas)	4	55,970	56,322	352	1%
Curragh (Derby Meeting)	3	10,022	12,821	2,799	28%
Fairyhouse (Easter)	2	22,560	16,002	-6,558	-29%
Limerick (Christmas)	4	40,072	40,426	354	1%
Total	**32**	**487,107**	**487,812**	**6,787**	**1%**



ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ