Πανεπιστήμιο Πειραιώς

**Τμήμα Ψηφιακών Συστημάτων**

ΠΜΣ «Πληροφοριακά Συστήματα & Υπηρεσίες»

Κατεύθυνση «Προηγμένα Πληροφοριακά Συστήματα»



Αποθήκες Δεδομένων και Επιχειρηματική Ευφυΐα

Απαλλακτική Εργασία 2020-2021

«Πρόβλεψη Ερευνητικών Τάσεων από Βιβλιογραφικά Δεδομένα»

Μεταπτυχιακός Φοιτητής

Αυγέρος Νικόλαος  
ΑΜ (ΜΕ2042)

Περίληψη

Η παρούσα εργασία έχει ως σκοπό την ανάπτυξη ενός μοντέλου πρόβλεψης της ερευνητικής τάσης για τα επόμενα δέκα χρόνια από τα βιβλιογραφικά δεδομένα του DBLP ιστότοπου, που διατηρεί βιβλιογραφικά στοιχεία δημοσιεύσεων στον χώρο της επιστήμης των υπολογιστών. Ειδικότερα, το μοντέλο που αναπτύχθηκε προβλέπει το πλήθος των δημοσιεύσεων για τη θεματική περιοχή της τεχνικής νοημοσύνης για τα επόμενα δέκα έτη. Για την επίτευξη αυτού, δημιουργήθηκε ένας text parser για την εξαγωγή συναφών δημοσιεύσεων, ο οποίος επιστρέφει το πλήθος των δημοσιεύσεων ανά έτος. Στη συνέχεια, εφαρμόστηκαν τα δεδομένα σε ένα πλαίσιο δεδομένων για τους σκοπούς της ανάλυσης και πρόβλεψης. Ακόμη, πραγματοποιήθηκε καθαρισμός δεδομένων από ελλιπής τιμές και στη συνέχεια χρησιμοποιήθηκαν τεχνικές οπτικοποίησης δεδομένων για την κατανόηση της δομής των δεδομένων. Επιπλέον, έγινε χρήση της τεχνική εξομάλυνσης Holt Winter και ακολούθησε η μετατροπή της χρονοσειράς σε στάσιμη. Τέλος, κατασκευάστηκαν μοντέλα πρόβλεψης ARIMA και Holt Winter και ακολούθησε σύγκριση των αποτελεσμάτων.

Για την υλοποίηση των ανωτέρω χρησιμοποιήθηκε η γλώσσα προγραμματισμού python σε περιβάλλον Jupyter, καθώς έγινε χρήση πληθώρας βιβλιοθηκών όπως Pandas, Numpy, Sklearn Matplotlib, Statsmodel, Math, Seaborn και Pylab.

**Περιεχόμενα**

[Βήμα 1: 4](#_Toc70860423)

[Βήμα 2: 4](#_Toc70860424)

[Βήμα 3: 5](#_Toc70860425)

[Βήμα 4: 8](#_Toc70860426)

## Βήμα 1:

Για τους σκοπούς της εργασίας χρησιμοποιήθηκε το αρχείο “dblp-2021-02-01.xml.gz” από τον ιστότοπο <http://dblp.org/xml/release/> μέσα από το οποίο προσφέρονται όλες οι απαραίτητες πληροφορίες για να πραγματοποιηθεί η ανάλυση χρονοσειρών για τη θεματική περιοχή της τεχνικής νοημοσύνης.

## Βήμα 2:

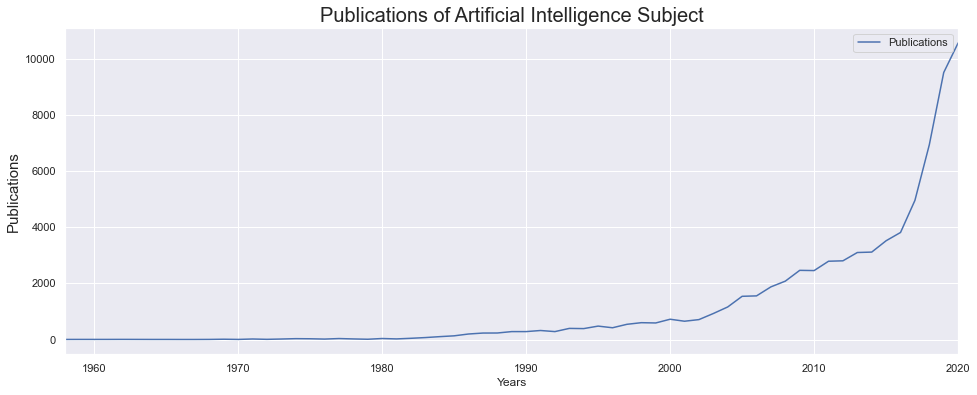
Για την εξαγωγή των δεδομένων από το αρχείο “dblp-2021-02-01.xml.gz” αναπτύχθηκε ένας text parser με τη γλώσσα προγραμματισμού Python. Ο συγκεκριμένος text parser, αφότου ανοίξει το αρχείο, δημιουργεί ένα λεξικό (year\_dictionary) με την εντολή dict, όπου αποθηκεύονται ως κλειδιά (keys) τα έτη (Years) που έχουν δημοσιευθεί θέματα σχετικά με την τεχνική νοημοσύνη και ως τιμές (values) το πλήθος αυτών, δηλαδή το πόσες φορές εμφανίζονται τα έτη στο συγκεκριμένο λεξικό.

Αυτό επιτυγχάνεται δημιουργώντας μια λίστα (subjects) που περιέχει το όνομα της θεματικής περιοχής που μας ενδιαφέρει, καθώς και συναφείς λέξεις. Οι λέξεις που χρησιμοποιήθηκαν ως λίστα για να εντοπιστούν οι δημοσιεύσεις είναι οι ακόλουθες: artificial intelligence, ΑΙ, advanced information system, artificial, intelligence, robotics, intelligent agent, computer interaction, machine learning, intelligent system, smart assistant, cognitive system.

Στη συνέχεια, ο parser διαβάζει το εν λόγω αρχείο γραμμή γραμμή και εντοπίζει πρώτα τους τίτλους που περιέχουν τις συναφείς λέξεις. Σημειώνεται ότι στον κώδικα έχει δημιουργηθεί προηγουμένως ένας μετρητής (number\_publication), ο οποίος αυξάνεται κατά μια μονάδα κάθε φορά που επιστρέφεται μια δημοσίευση. Σκοπός αυτού είναι για να εκτυπώνεται στην κονσόλα τόσο ο αριθμός της δημοσίευσης όσο και ο συνολικός αριθμός τους, ώστε να είναι ευδιάκριτα τα αποτελέσματα κατά την εκτέλεση του προγράμματος. Ακόμη, για να εντοπιστούν αποτελεσματικά οι τίτλοι και τα έτη που περιέχονται στο αρχείο xml, χρησιμοποιούνται δομές ελέγχου για τον εντοπισμό των λέξεων <title> και <year>.

Προκειμένου να μην υπάρξει απώλεια από τους τίτλους των δημοσιεύσεων κατά την εξαγωγή τους χρησιμοποιήθηκε η εντολή lower, η οποία μετατρέπει τα κεφαλαία σε πεζά εάν υπάρχουν. Εν συνεχεία, εφόσον εντοπίσει ο parser τον τίτλο της δημοσίευσης, συνεχίζει να διαβάζει την αμέσως επόμενη γραμμή με την εντολή readline για να ελεγχθεί εάν πρόκειται για έτος. Στην περίπτωση που δεν αντιστοιχεί η επόμενη γραμμή σε κάποιο έτος, τότε συνεχίζει ξανά με την εντολή readline στην ακριβώς επόμενη γραμμή. Η διαδικασία αυτή επαναλαμβάνεται τέσσερις φορές στον κώδικα, που σημαίνει ότι εφόσον εντοπίσει τον τίτλο (<title>) θα διαβάσει τις επόμενες τέσσερις γραμμές για να εντοπίσει το έτος. Στο τέλος του parser και για την αποφυγή διπλότυπων δημοσιεύσεων δίδεται η εντολή break για να βγεί από την επανάληψη. Τέλος, τα δεδομένα ταξινομούνται σε δύο στήλες σε έτη και δημοσιεύσεις κατά αύξουσα σειρά και αποθηκεύονται σε ένα αρχείο τύπου κειμένου (text).

Στο διάγραμμα που ακολουθεί απεικονίζονται τα δεδομένα των δημοσιεύσεων ανά έτος, όπως αυτές εξάχθηκαν από τον parser που περιεγράφηκε ανωτέρω και αφού μετατράπηκαν σε πλαίσιο δεδομένων (dataframe) με τη χρήση της βιβλιοθήκης pandas. Για την οπτικοποίηση των δεδομένων χρησιμοποιήθηκε η βιβλιοθήκη matplotlib.



Εικόνα 1: Απεικόνιση των δημοσιεύσεων ανά έτος

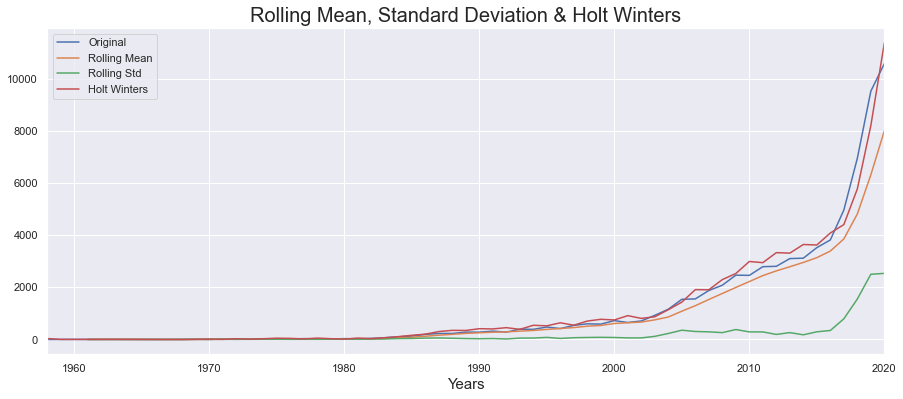
Σύμφωνα με το διάγραμμα παρατηρούμαι ότι η χρονοσειρά ξεκινάει από το έτος 1958 και τελειώνει έως το 2020, καθώς επίσης τα δεδομένα μας αυξάνονται εκθετικά με την πάροδο του χρόνου ξεπερνώντας τις δέκα χιλιάδες δημοσιεύσεις σε θέματα που αφορούν στην τεχνητή νοημοσύνη το 2020. Επομένως, καταλήγουμε στο συμπέρασμα ότι η χρονοσειρά μας εμφανίζει τάση, ενώ δεν έχει τα χαρακτηριστικά της εποχικότητας, της κυκλικότητας και των ακραίων τιμών.

Αξίζει να σημειωθεί ότι για την αντιμετώπιση των ελλιπών τιμών χρησιμοποιήθηκε η βιβλιοθήκη pandas. Ειδικότερα, αφαιρέθηκαν τα έτη πριν το 1958, διότι είναι αρκετά παρελθοντικά με τιμές του πλήθους των δημοσιεύσεων να αντιστοιχεί σε τιμές μικρότερες των δυο. Επομένως κρίθηκαν περιττές και δεν θα είχαν ιδιαίτερη αξία στην αντιμετώπιση αντικατάστασης τους. Αντιθέτως, τα έτη 1960 και 1965 αντικαταστάθηκαν με τις τιμές των αμέσως επόμενων ετών, με τη χρήση της μεθόδου ffill.

## Βήμα 3:

Αρχικά, για να επιβεβαιώσουμε τις παρατηρήσεις της ανωτέρω περιγραφής του διαγράμματος θα πρέπει να ελέγξουμε τη στασιμότητα της χρονοσειράς. Για τον έλεγχο της στασιμότητας χρησιμοποιήθηκε η διπλή εκθετική εξομάλυνση που βοηθάει στην απεικόνιση των αποτελεσμάτων και η στατιστική μέθοδος Dickey-Fuller test.

Στο διάγραμμα που ακολουθεί απεικονίζουμε εκτός από τον κινητό μέσο και τη μέθοδο Holt Winter, καθώς ο κινητός μέσος παρουσιάζει αρκετά μειονεκτήματα, όπως ότι δεν μπορεί να εφαρμοστεί στα άκρα του συνόλου δεδομένων, είναι ακριβός υπολογιστικά και δεν ορίζεται εκτός του εύρους του πεδίου ορισμού γεγονός που δεν χρησιμεύει για την πρόβλεψη. Για τη κατασκευή του διαγράμματος χρησιμοποιήθηκε διάστημα τεσσάρων ετών (window = 4), ενώ για την εφαρμογή του Holt Winter ορίσαμε παράμετρο την τάση ως πολλαπλασιαστική (trend = ‘mull’), αφού πρόκειται για εκθετική αύξηση της τάση. Σημειώνεται ότι χρησιμοποιείται η διπλή εκθετική εξομάλυνση, καθώς όπως παρατηρήθηκε η χρονοσειρά μας παρουσιάζει μόνο τάση και καθόλου εποχικότητα.



Εικόνα : Απεικόνιση τεχνικών εξομάλυνσης Κινητού μέσου & Holt Winters.

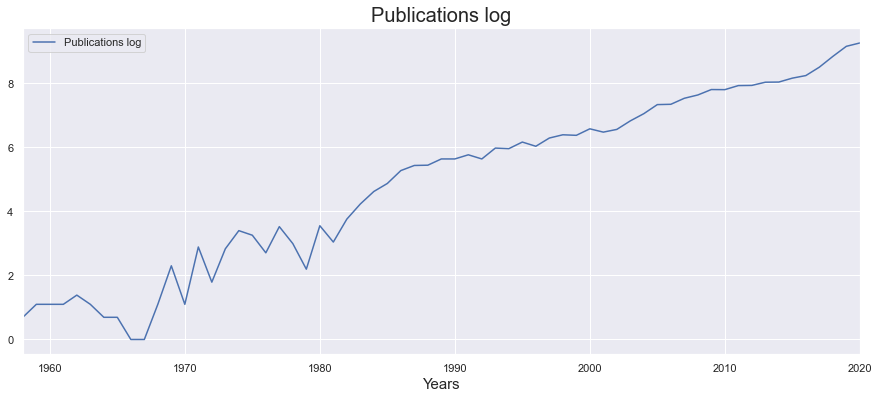
Από την ανάλυση του διαγράμματος παρατηρούμε ότι ο κινητός μέσος και ο Holt Winters αυξάνονται εκθετικά, ενώ η τυπική απόκλιση βρίσκεται σε χαμηλά επίπεδα και δεν είναι σταθερή γεγονός που οδηγεί στο συμπέρασμα ότι η χρονοσειρά είναι μη στάσιμη.

Συνεχίζοντας με τον έλεγχο στασιμότητας της χρονοσειράς διεξάγουμε το Dickey Fuller test χρησιμοποιώντας τη βιβλιοθήκη statsmodels (adfuller).

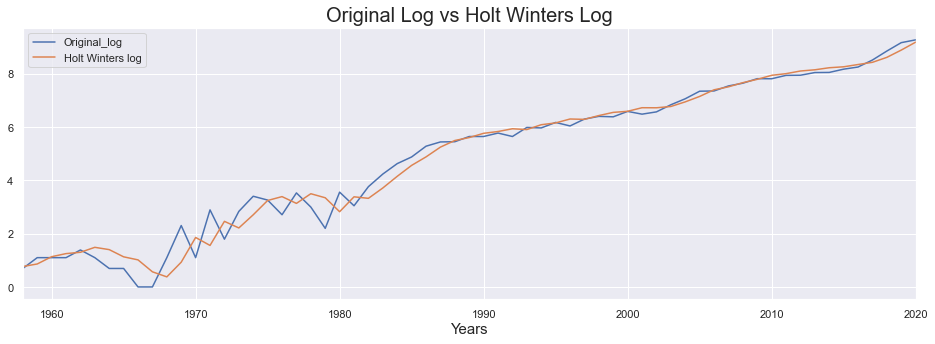
|  |  |
| --- | --- |
| Dickey Fuller test | |
| ADFuller Statistics | 4.239072527128941 |
| P value: | 1.0 |
| Lags: | 11 |
| Cristical values |  |
| 1% | -3.566 |
| 5% | -2.920 |
| 10% | -2.598 |
|  |  |

Από τα αποτελέσματα διακρίνουμε ότι ο test statistic είναι αρκετά μεγαλύτερος από τα επίπεδα σημαντικότητας (critical values) και η p value είναι αρκετά μεγάλη. Αυτό που μας αποδεικνύει ότι δεν μπορούμε να απορρίψουμε την H0 που σημαίνει ότι η χρονοσειρά δεν είναι στάσιμη. Για να απορρίψουμε την H0 και να δεχτούμε την εναλλακτική υπόθεση πρέπει η Test Statistic να είναι μικρότερη από τις critical values και η P value να είναι τουλάχιστον 0,5 ή μικρότερη.

Για να προχωρήσουμε στην κατασκευή ενός μοντέλου πρόβλεψης πρέπει η χρονοσειρά μας να γίνει στάσιμη, δηλαδή να αφαιρέσουμε την τάση και την εποχικότητα. Επομένως, πρέπει να εκτιμήσουμε την τάση μετασχηματίζοντας αρχικά τη χρονοσειρά σε χαμηλότερες τιμές με τη χρήση λογαριθμικού, χρησιμοποιώντας τη βιβλιοθήκη numpy και στη συνέχεια θα χρησιμοποιήσουμε τη μέθοδο εξομάλυνσης Holt Winters.

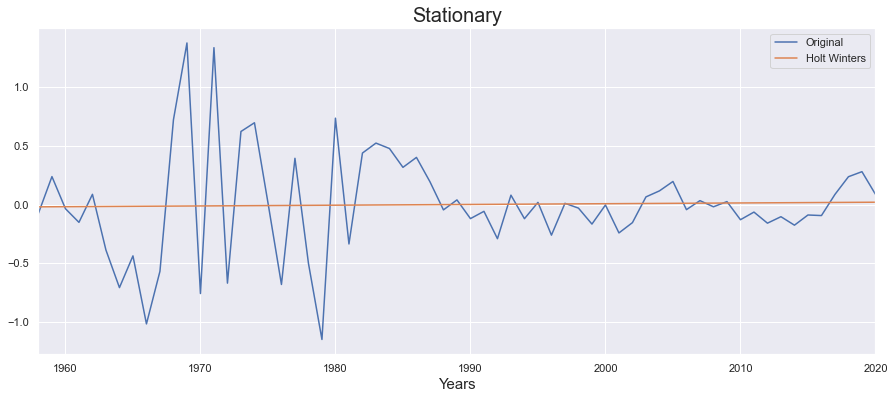


Από το διάγραμμα παρατηρούμε ότι οι τιμές των δημοσιεύσεων έχουν μειωθεί σημαντικά και η τάση δείχνει να αυξάνεται γραμμικά αυτή τη φορά. Στη συνέχεια θα εφαρμόσουμε και θα παρουσιάσουμε διαγραμματικά τη μέθοδο Holt Winters στα λογαριθμικά δεδομένα μας και θα τα συγκρίνουμε.



Από τα αποτελέσματα του διαγράμματος είναι εμφανές ότι η χρονοσειρά μας έχει εξομαλυνθεί και μοιάζει περισσότερο γραμμική. Η χρονοσειρά μας είναι πλέον έτοιμη για πρόβλεψη με τη μέθοδο Holt Winter, όμως επειδή για το λόγο ότι θα χρησιμοποιήσουμε και μοντέλο ARIMA για πρόβλεψη θα πρέπει να μετατρέψουμε τη χρονοσειρά σε στάσιμη.

Προκειμένου να την κάνουμε στάσιμη θα αφαιρέσουμε τον Holt Winter από τη χρονοσειρά μας, αλλά από τα μετασχηματισμένα δεδομένα.

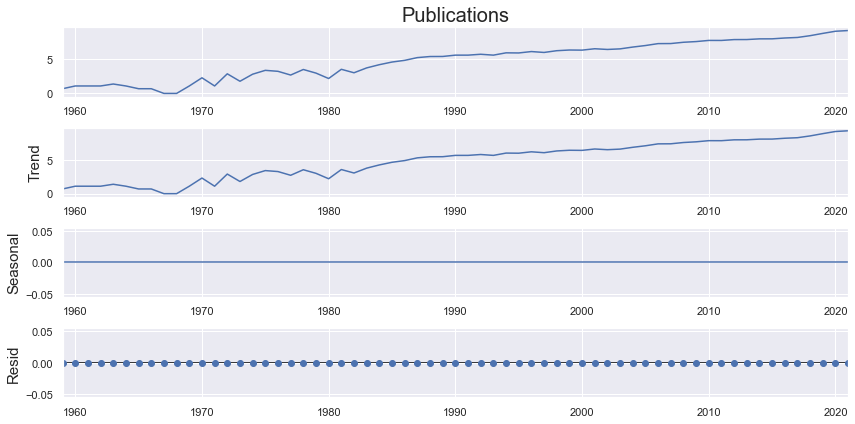


Είναι εμφανές ότι η χρονοσειράς μας δεν εμφανίζει πλέον καμία τάση και δείχνει να είναι στάσιμη. Για να επιβεβαιώσουμε τη στασιμότητα θα πραγματοποιήσουμε έλεγχο στασιμότητας κάνοντας Dickey Fuller test.

|  |  |
| --- | --- |
| Dickey Fuller test | |
| ADFuller Statistics | -4.998431980346559 |
| P value: | 2.235228433534205e-05 |
| Lags: | 7 |
| Cristical values |  |
| 1% | -3.555 |
| 5% | -2.916 |
| 10% | -2.596 |
|  |  |

Από τον στατιστικό έλεγχο παρατηρούμε ότι η p value είναι πολύ μικρή και ο test statistic είναι μικρότερος από όλα τα επίπεδα σημαντικότητας, άρα μπορούμε να πούμε ότι είμαστε 100% σίγουροι ότι η χρονοσειρά μας είναι στάσιμη.

Μπορούμε βέβαια να ξεχωρίσουμε με εύκολο τρόπο τα διάφορα χαρακτηριστικά τηε χρονοσειράς μας, όπως την τάση, την εποχικότητα και τα σφάλματα με τη χρήση της βιβλιοθήκης statsmodels (seasonal\_decompose).



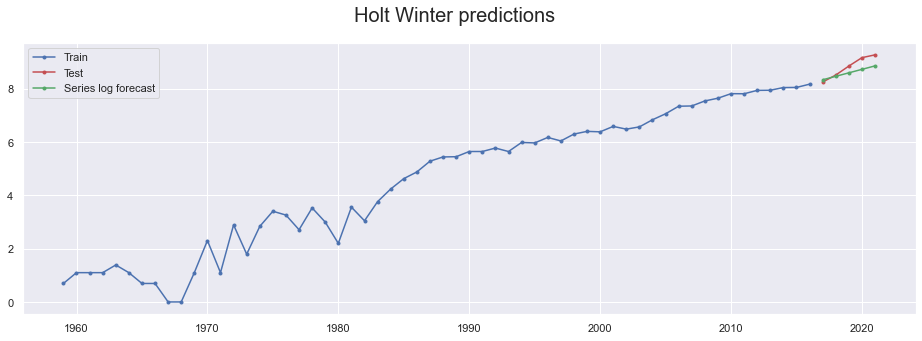
Στο παραπάνω διάγραμμα φαίνονται τα αποτελέσματα και είναι εμφανές ότι η χρονοσειρά εμφανίζει τάση, ενώ απουσιάζουν η εποχικότητα και τα σφάλματα και επομένως η χρονοσειρά μας είναι έτοιμη για το μοντέλο πρόβλεψης που θα κατασκευάσουμε.

## Βήμα 4:

Η μέθοδοι που θα χρησιμοποιήσουμε για την πρόβλεψη της χρονοσειράς είναι η μέθοδος ARIMA και μέθοδος Holt Winter. Αρχικά, χωρίζουμε τα δεδομένα μας σε train και test , ώστε να χρησιμοποιήσουμε τα δεδομένα ως το 2015 και να γίνει έλεγχος της ακρίβειας πρόβλεψης για τα έτη 2016 έως 2020 που υπάρχει διαθέσιμο το πραγματικό πλήθος δημοσιεύσεων. Εφαρμόζοντας με το μοντέλο πρόβλεψης Holt Winter έχουμε τα εξής αποτελέσματα:

|  |  |  |
| --- | --- | --- |
| **Στοιχεία πρόβλεψης** | Holt Winter | Test |
| 2016 | 4153.799222 | 3818 |
| 2017 | 7431.914937 | 4963 |
| 2018 | 5390.491396 | 6948 |
| 2019 | 6140.726931 | 9524 |
| 2020 | 6995.378432 | 10578 |

Από τα ανωτέρω διακρίνουμε ότι υπάρχει σημαντική απόκλιση στην ακρίβεια από τα δεδομένα πρόβλεψης σε σύγκριση με τα πραγματικά δεδομένα ελέγχου. Παρόλα αυτά το μοντέλο πρόβλεψης που κατασκευάσαμε προβλέπει την ανοδική πορεία που έχουν τα πραγματικά δεδομένα ελέγχου. Αυτό μπορούμε να το διακρίνουμε γρήγορα και διαγραμματικά.



Από το διάγραμμα επιβεβαιώνουμε την ανοδική πορεία δεδομένων πρόβλεψης, ενώ διακρίνουμε την απόκλιση που έχουν οι δύο ευθείας μεταξύ τους. Για την αξιολόγηση της επίδοσης του μοντέλου πρόβλεψης θα χρησιμοποιήσουμε τα στατιστικά μέτρα, όπως το μέσο τετραγωνικό σφάλμα (MSE), τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) και μέσο απόλυτο σφάλμα (MAE).

Τα αποτελέσματα της εκτέλεσης του κώδικα έχουμε:

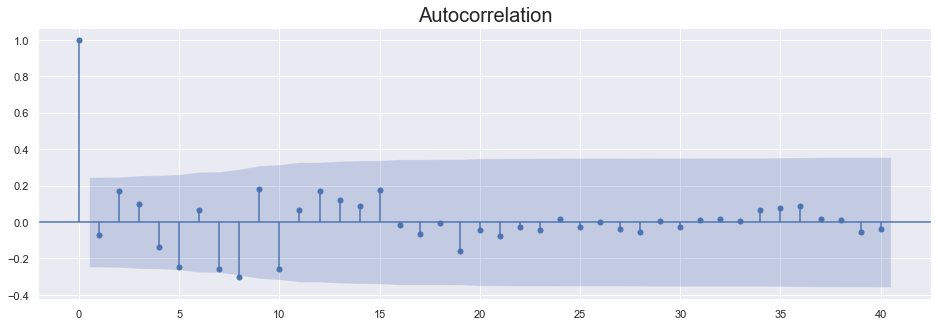
|  |  |
| --- | --- |
| **Αξιολόγηση επίδοσης του μοντέλου πρόβλεψης Holt Winter** | |
| Root Mean Squared Error (RMSE) | 0.2957753651659285 |
| Mean Absolute Error (MAE) | 0.24763857165004097 |
| Mean Squared Error (MSE) | 0.087 |

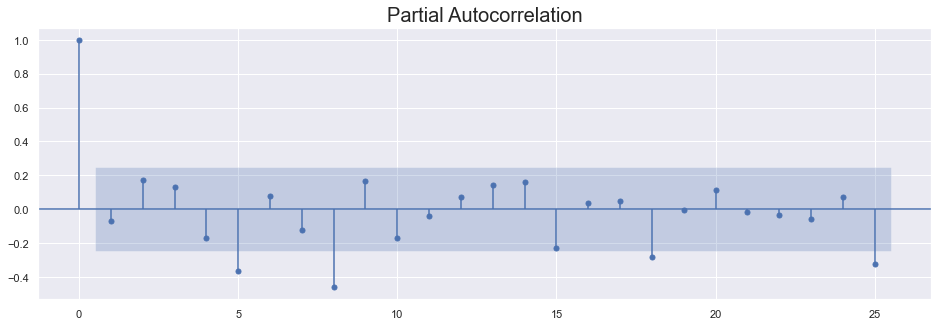
Από τα αποτελέσματα διακρίνουμε ότι οι τιμές των δεικτών είναι σχετικά μεγάλοι για να προτιμήσουμε το συγκεκριμένο μοντέλο.

Για την κατασκευή του μοντέλου ARIMA απαιτούνται τρεις βασικές παράμετροι:

1. **Ο αριθμός των διαφορών (d):** Αφορά στον αριθμό των διαφορών που πραγματοποιήσαμε για τη μετατροπή της χρονοσειράς σε στάσιμη.
2. **Αυτοπαλινδρομούμενη χρονοσειρά (autoregressive time series τάξης p (AR(p)):** Αφορά στις χρονικές υστερήσεις της εξαρτημένης μεταβλητής.
3. **Χρονοσειρά κινητού μέσου τάξης q (MA(q)):** Αφορά στις χρονικές υστερήσεις των σφαλμάτων πρόβλεψης.

Από τα παραπάνω γνωρίζουμε τη παράμετρο d και πρέπει να καθορίσουμε τις παραμέτρους p και q για την κατασκευή του μοντέλου πρόβλεψης. Αυτό επιτυγχάνεται με την αυτοσυσχέτιση και τη βοήθεια των διαγραμμάτων ACF και PACF.

****

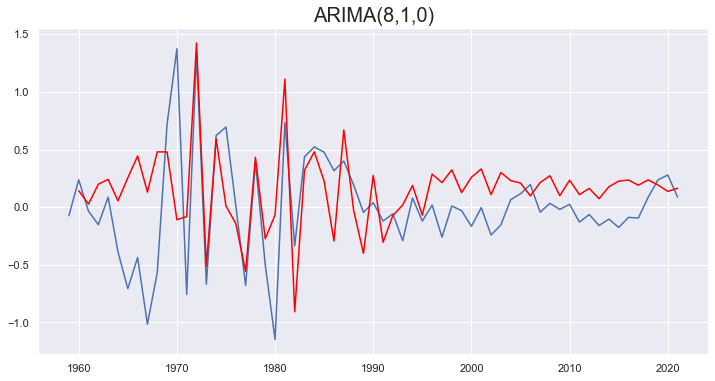
****

Από τα διαγράμματα συμπεραίνουμε ότι η τιμή p ισούται με 8 και η τιμή q με 1. Όμως, η μελέτη των διαγραμμάτων αποτελεί δύσκολη μέθοδο για την εύρεση των κατάλληλων παραμέτρων, ώστε να κατασκευαστεί το μοντέλο ARIMA. Χάρη στην υπολογιστική ισχύ μπορούμε να υλοποιήσουμε τη διαδικασία cross validation, όπου τα αποτελέσματα της διασταυρούμενης επικύρωσης μπορούν να χρησιμοποιηθούν για την επιλογή του καλύτερου συνόλου τιμών παραμέτρων του μοντέλου ARIMA. Τα αποτελέσματα που προκύπτουν από την εκτέλεση του κώδικα είναι τα ακόλουθα:

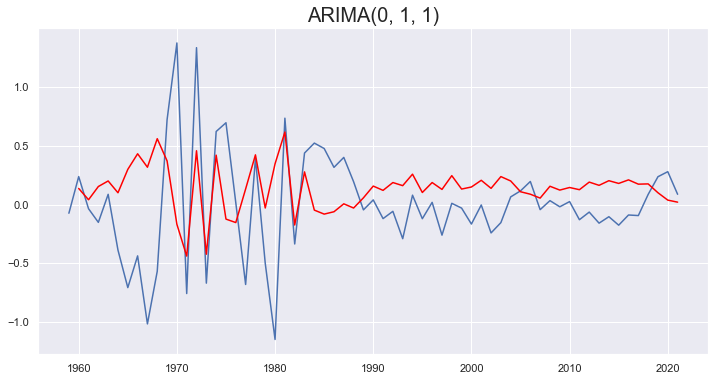
|  |  |
| --- | --- |
| Cross - validation | |
| ARIMA(0, 0, 0) | MSE=17.192 |
| ARIMA(0, 0, 1) | MSE=5.330 |
| ARIMA(0, 1, 0) | MSE=0.019 |
| ARIMA(0, 1, 1) | MSE=0.033 |
| ARIMA(4, 0, 0) | MSE=0.096 |
| ARIMA(4, 0, 1) | MSE=0.057 |
| ARIMA(4, 1, 0) | MSE=0.037 |
| ARIMA(4, 1, 1) | MSE=0.043 |
| ARIMA(5, 0, 0) | MSE=0.098 |
| ARIMA(5, 0, 1) | MSE=0.095 |
| ARIMA(5, 1, 0) | MSE=0.026 |
| ARIMA(5, 1, 1) | MSE=0.016 |
| ARIMA(8, 0, 0) | MSE=0.088 |
| ARIMA(8, 0, 1) | MSE=0.081 |
| ARIMA(8, 1, 0) | MSE=0.014 |
| ARIMA(8, 1, 1) | MSE=0.013 |
| **Best ARIMA(8, 1, 1)** | **MSE=0.013** |

Από τα αποτελέσματα δείχνουν ότι το βέλτιστο μοντέλο ARIMA έχει τιμές 8, 1, 1 για τις παραμέτρους q, d, p αντίστοιχα. Επιπλέον, το μέσο τετραγωνικό σφάλμα έχει αρκετά χαμηλότερη τιμή συγκριτικά με την προηγούμενη πρόβλεψη του Holt Winter που σημαίνει ότι το συγκεκριμένο μοντέλο είναι προτιμότερο. Στη συνέχεια, θα δούμε διαγραμματικά το μοντέλο ARIMA(8,1,1) και θα το συγκρίνουμε με τα μοντέλα AR και MA

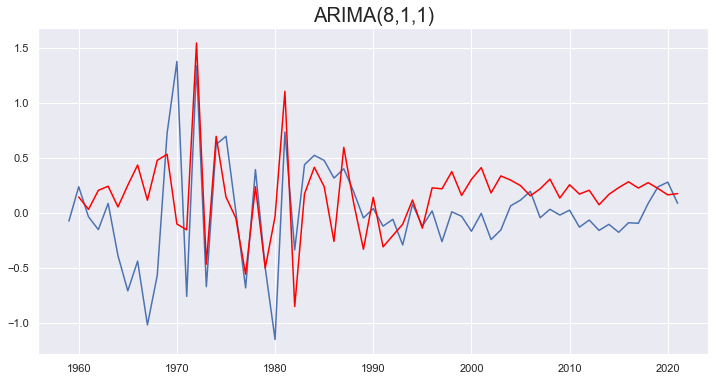
AR model:



MA model:



Best ARIMA model:



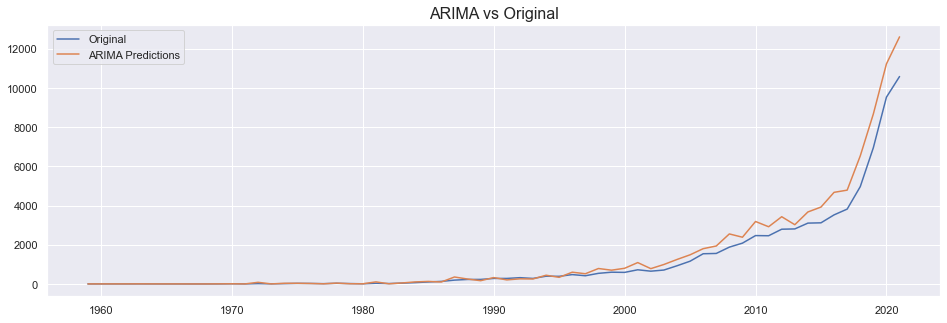
Συνεχίζοντας, εφόσον βρήκαμε ποιο μοντέλο ARIMA είναι το πιο κατάλληλο το εφαρμόζουμε στο σύνολο δεδομένων προς εκπαίδευση και συγκρίνουμε τα αποτελέσματα με το σύνολο δεδομένων για έλεγχο. Από την εκτέλεση έχουμε τα εξής αποτελέσματα:

|  |  |  |
| --- | --- | --- |
| **Στοιχεία πρόβλεψης** | ARIMA (8,1,1) | Test |
| 2016 | 4784.266278 | 3818 |
| 2017 | 6531.226708 | 4963 |
| 2018 | 8669.407187 | 6948 |
| 2019 | 11220.671554 | 9524 |
| 2020 | 12601.014488 | 10578 |

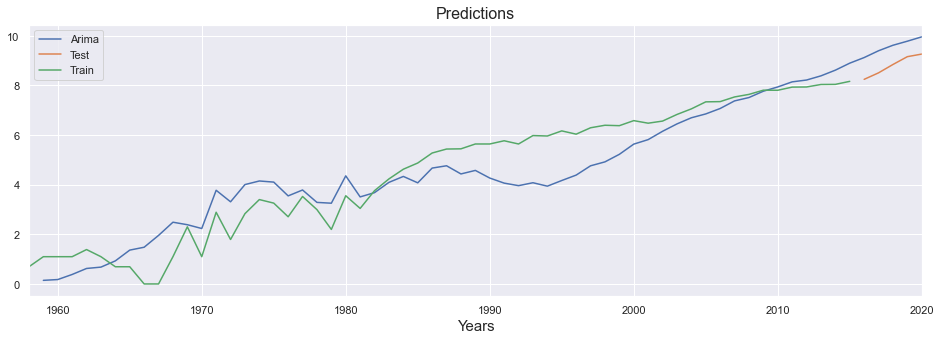
Από τα αποτελέσματα βλέπουμε ότι εξακολουθούμε να έχουμε κάποια απόκλιση από τα δεδομένα ελέγχου, αλλά και σε αυτή την περίπτωση τα δεδομένα πρόβλεψης αυξάνονται εκθετικά. Υπολογίζοντας τα σφάλματα έχουμε τα εξής αποτελέσματα:

|  |  |
| --- | --- |
| **Αξιολόγηση επίδοσης του μοντέλου πρόβλεψης ARIMA(8, 1, 1)** | |
| Root Mean Squared Error (RMSE) | 0.11393886093060626 |
| Mean Absolute Error (MAE) | 0.09952643254152527 |
| Mean Squared Error (MSE) | 0.013 |

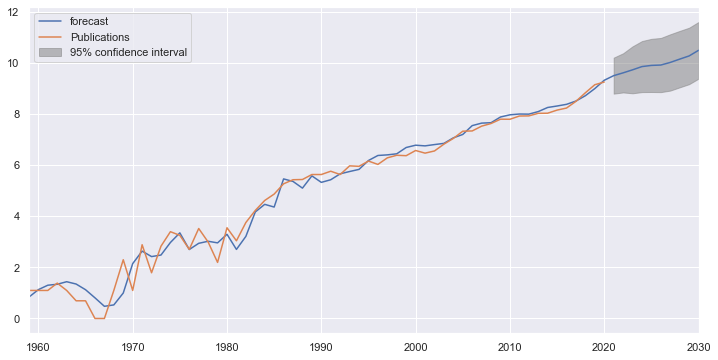
Συγκρίνοντας τα αποτελέσματα και οπτικά:



Είναι εμφανές το μοντέλο που κατασκευάσαμε προβλέπει τόσο την ανοδική πορεία των πραγματικών δεδομένων όσο και την εκθετική τους αύξηση. Διακρίνουμε ότι η πορτοκαλί γραμμή που αντιστοιχεί στο μοντέλο πρόβλεψης ακολουθάει τα πραγματικά δεδομένα μας που εμφανίζονται με τη μπλε γραμμή στο διάγραμμα που κατασκευάσαμε. Στο διάγραμμα που ακολουθεί βλέπουμε τα ίδια αποτελέσματα και στα λογαριθμικά δεδομένα μας



Στο τέλος απεικονίζουμε τα αποτελέσματα πρόβλεψης για τα επόμενα δέκα έτη.



Παρατηρούμε ότι η θεματική περιοχή της τεχνητή νοημοσύνης πρόκειται να σημειώσει ανοδική πορεία για τα επόμενα δέκα έτη και αυτό είναι ορθολογικό, καθώς η τεχνολογία εξελίσσεται ραγδαία τα τελευταία χρόνια με την τεχνητή νοημοσύνη να πρωταγωνιστεί στη ζωή μας. Το μοντέλο ARIMA(8,1,1) που κατασκευάσαμε δείχνει να ανταποκρίνεται σχετικά καλά συγκριτικά με τα δεδομένα μας.