ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Μέθοδοι Πρόβλεψης σε Χρονοσειρές Μετεωρολογικών Μετρήσεων

Michalaki Elisavet

University of Piraeus

MSc in Cyber Security and Business Analytics

Time-Series Analytics and Forecasting

Athens, January 2025

Table of Contents

Αρχική επεξεργασία των δεδομένων..... 3
Autoregressive Μοντέλο..... 5
Νευρωνικά δίκτυα..... 7

Αρχική επεξεργασία των δεδομένων

Η πηγή των δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία είναι ένα αρχείο .CSV με τίτλο cleaned_weather.csv, το οποίο περιλαμβάνει πλήθος μετεωρολογικών μετρήσεων καταγεγραμμένων σε χρονική σειρά, μεταξύ άλλων, τη θερμοκρασία (T), το σημείο δρόσου (Tdew), την πραγματική τάση υδρατμών (VPact), την εν δυνάμει θερμοκρασία (Tpot) και την ηλιακή ακτινοβολία (PAR).

	date	p	T	Tpot	Tdew	rh	VPmax	VPact	\
0	2020-01-01 00:10:00	1008.89	0.71	273.18	-1.33	86.1	6.43	5.54	
1	2020-01-01 00:20:00	1008.76	0.75	273.22	-1.44	85.2	6.45	5.49	
2	2020-01-01 00:30:00	1008.66	0.73	273.21	-1.48	85.1	6.44	5.48	
3	2020-01-01 00:40:00	1008.64	0.37	272.86	-1.64	86.3	6.27	5.41	
4	2020-01-01 00:50:00	1008.61	0.33	272.82	-1.50	87.4	6.26	5.47	

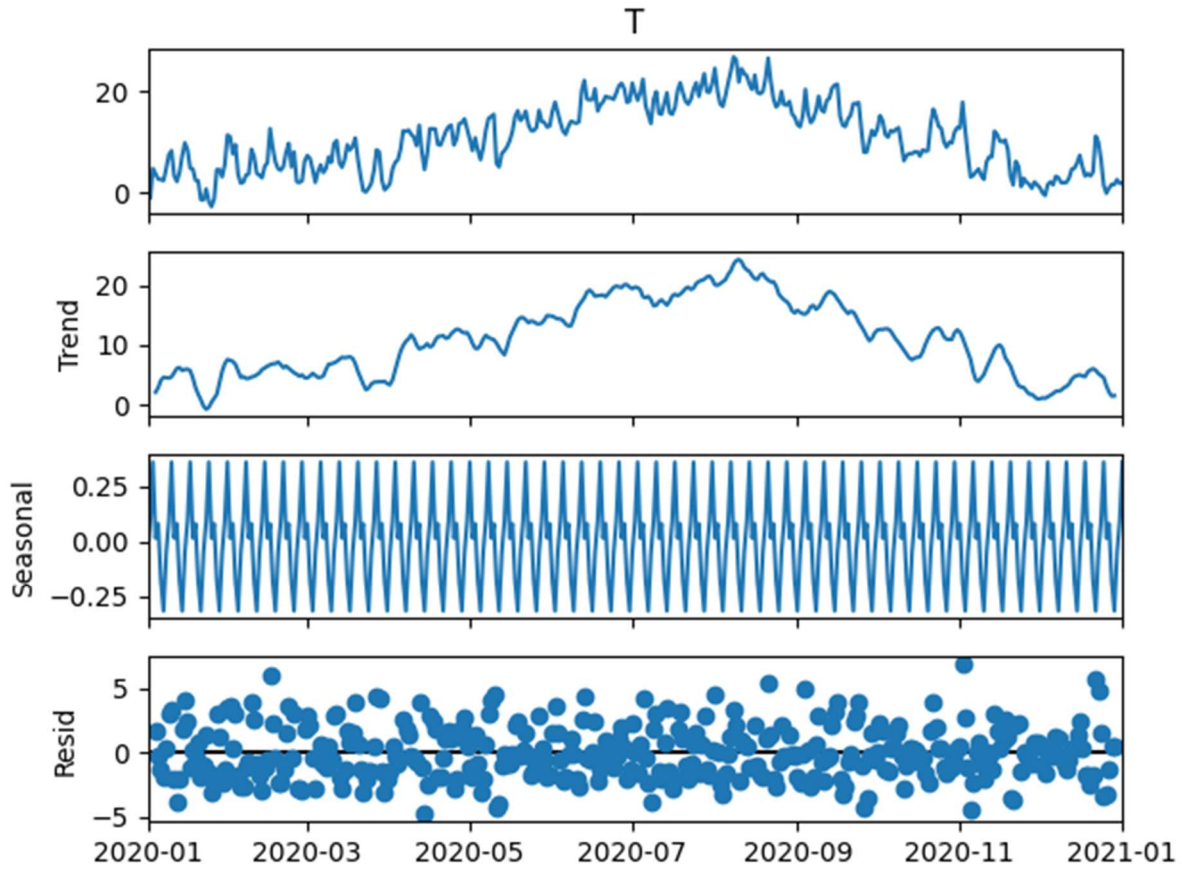
	VPdef	sh	...	rho	wv	max. wv	wd	rain	raining	SWDR	PAR	\
0	0.89	3.42	...	1280.62	1.02	1.60	224.3	0.0	0.0	0.0	0.0	
1	0.95	3.39	...	1280.33	0.43	0.84	206.8	0.0	0.0	0.0	0.0	
2	0.96	3.39	...	1280.29	0.61	1.48	197.1	0.0	0.0	0.0	0.0	
3	0.86	3.35	...	1281.97	1.11	1.48	206.4	0.0	0.0	0.0	0.0	
4	0.79	3.38	...	1282.08	0.49	1.40	209.6	0.0	0.0	0.0	0.0	

	max. PAR	Tlog
0	0.0	11.45
1	0.0	11.51
2	0.0	11.60
3	0.0	11.70
4	0.0	11.81

[5 rows x 21 columns]

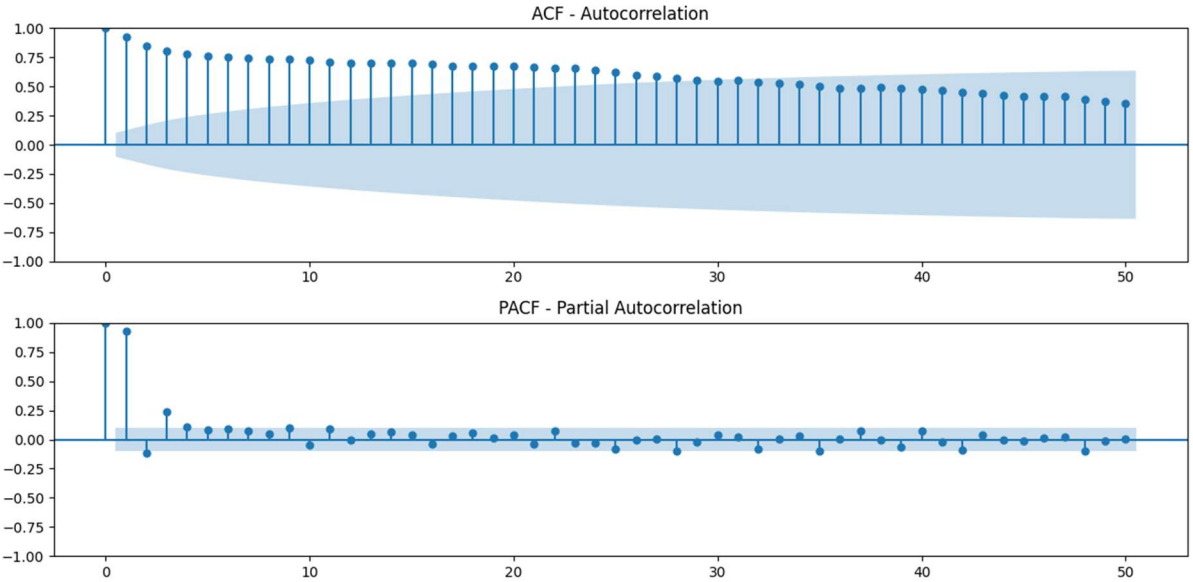
Αρχικά, η στήλη date, η οποία περιείχε πληροφορίες για την ημερομηνία και την ώρα των καταγραφών, μετατράπηκε σε τύπο datetime. Στη συνέχεια, η στήλη αυτή ορίστηκε ως index του DataFrame. Εφαρμόστηκε επαναδειγματοληψία σε ημερήσια βάση, κατά την οποία για κάθε ημέρα υπολογίστηκε ο μέσος όρος των τιμών, με στόχο την εξομάλυνση των υψηλών διακυμάνσεων και την αποφυγή υπερπληροφόρησης που μπορεί να οδηγήσει σε overfitting του μοντέλου.

Ακολούθησε ανάλυση των δεδομένων, η οποία περιλάμβανε την αποσύνθεση της χρονοσειράς της θερμοκρασίας με σκοπό την ανάδειξη της υποκείμενης τάσης, της εποχικότητας κλπ.

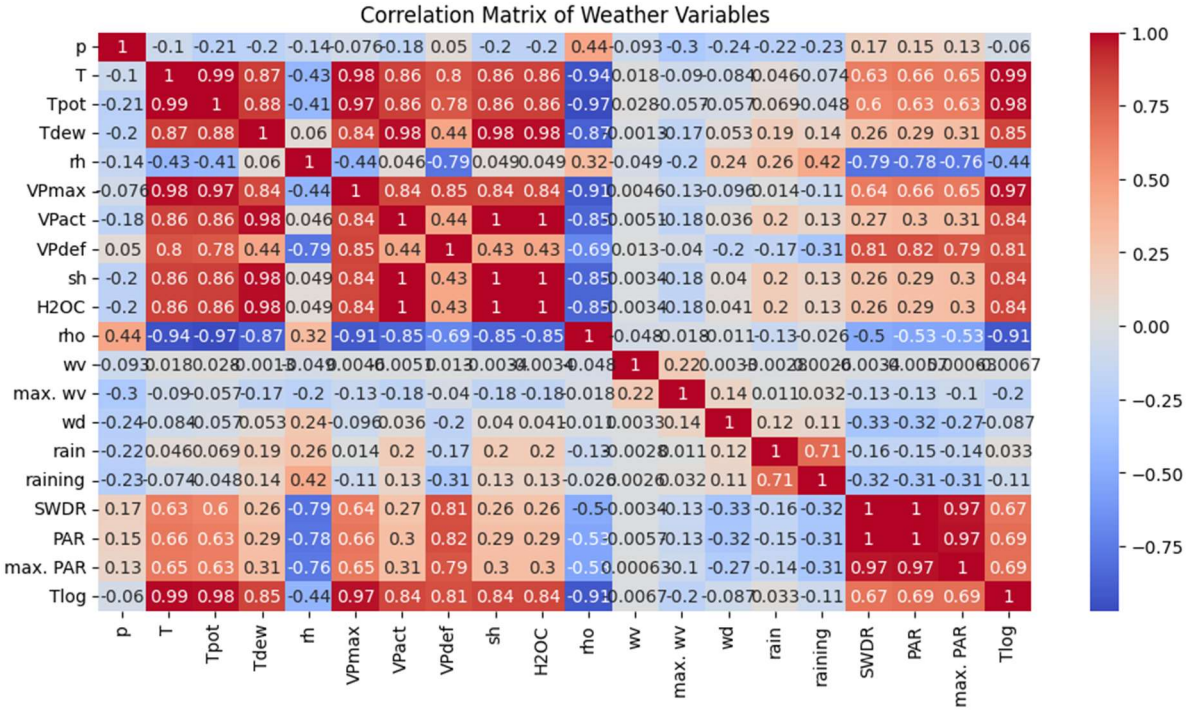


Το μοντέλο αποσύνθεσης ανέδειξε έντονη εβδομαδιαία περιοδικότητα και μια σταθερή μακροχρόνια τάση, γεγονός που ενισχύει την επιλογή της μεταβλητής θερμοκρασίας ως στόχο πρόβλεψης.

Επιπλέον, χρησιμοποιήθηκαν τα διαγράμματα αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) για την ανάλυση της εσωτερικής χρονικής εξάρτησης της χρονοσειράς.



Στη συνέχεια δημιουργήθηκε ένα heatmap όπου ανέδειξε ισχυρές συσχετίσεις μεταξύ της θερμοκρασίας και άλλων μεταβλητών όπως το Tpot, το Tdew και η VPact.



Αποφασίστηκε πως θα χρησιμοποιήσουμε την μεταβλητή T για πρόβλεψη γιατι είναι η πιο βασική και σημαντική μετεωρολογική μεταβλητή η οποία επηρεάζει σχεδόν όλα τα καιρικά φαινόμενα (π.χ. υγρασία, βροχή, εξάτμιση), έχει ισχυρές συσχετίσεις με πολλές άλλες μεταβλητές στο dataset (π.χ. Tpot, Tdew, VPact, PAR), παρουσιάζει εποχικότητα και τάση και είναι εύκολη στην αξιολόγηση, καθώς είναι σε συνεχείς τιμές και μπορούμε να συγκρίνουμε εύκολα MAE, RMSE, MAPE μεταξύ των μοντέλων.

Autoregressive Μοντέλο

Αρχικά υλοποιήθηκε ένα autoregressive μοντέλο ως με lag 7 (μια εβδομάδα), το οποίο εκπαιδεύτηκε στο 80% του συνόλου των δεδομένων, ενώ το υπόλοιπο 20% χρησιμοποιήθηκε ως test set.

AutoReg Model Results

Dep. Variable:	T	No. Observations:	293
Model:	AutoReg(7)	Log Likelihood	-637.268
Method:	Conditional MLE	S.D. of innovations	2.246
Date:	Tue, 17 Jun 2025	AIC	1292.535
Time:	16:52:05	BIC	1325.439
Sample:	01-08-2020 - 10-19-2020	HQIC	1305.724

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.5560	0.298	1.868	0.062	-0.027	1.139
T.L1	0.9754	0.059	16.589	0.000	0.860	1.091
T.L2	-0.3054	0.082	-3.707	0.000	-0.467	-0.144
T.L3	0.0521	0.084	0.617	0.537	-0.113	0.218
T.L4	0.1199	0.084	1.425	0.154	-0.045	0.285
T.L5	-0.0025	0.084	-0.030	0.976	-0.167	0.162
T.L6	0.0137	0.081	0.168	0.866	-0.146	0.173
T.L7	0.1045	0.058	1.803	0.071	-0.009	0.218

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0227	-0.0000j	1.0227	-0.0000
AR.2	0.8991	-0.9345j	1.2968	-0.1281
AR.3	0.8991	+0.9345j	1.2968	0.1281
AR.4	-0.0012	-1.4214j	1.4214	-0.2501
AR.5	-0.0012	+1.4214j	1.4214	0.2501
AR.6	-1.4748	-0.7617j	1.6599	-0.4241
AR.7	-1.4748	+0.7617j	1.6599	0.4241

Τα αποτελέσματα του μοντέλου AutoReg(7) δείχνουν ότι η μεταβλητή θερμοκρασίας επηρεάζεται σημαντικά από τις δύο πρώτες χρονικές καθυστερήσεις, με τον συντελεστή L1 να είναι ιδιαίτερα υψηλός και στατιστικά σημαντικός ($p < 0.001$), γεγονός που επιβεβαιώνει την έντονη αυτοσυσχέτιση της θερμοκρασίας με την τιμή της την προηγούμενη ημέρα. Η L2 έχει επίσης στατιστικά σημαντική επίδραση, αλλά με αρνητικό πρόσημο. Η έβδομη υστέρηση L7, αν και οριακά σημαντική ($p \approx 0.071$), υποδηλώνει την παρουσία εβδομαδιαίας εποχικότητας στο μοτίβο. Επιπλέον, η ανάλυση των ριζών του πολυωνύμου του μοντέλου επιβεβαιώνει τη σταθερότητα του συστήματος, καθώς όλες οι ρίζες έχουν μέτρο μεγαλύτερο της μονάδας. Η ύπαρξη φανταστικών ριζών υποδηλώνει κυκλικά πρότυπα στη χρονοσειρά, ενισχύοντας την ένδειξη εποχικότητας. Συνολικά, το μοντέλο εμφανίζεται σταθερό και ερμηνεύσιμο, με επαρκή δυναμική για την αποτύπωση βραχυπρόθεσμων και περιοδικών μεταβολών της θερμοκρασίας.

Το μοντέλο AR είναι χρήσιμο ως baseline, αλλά δεν συλλαμβάνει μη γραμμικές σχέσεις και δεν αξιοποιεί πολλαπλές μεταβλητές ή μακροχρόνιες εξαρτήσεις.

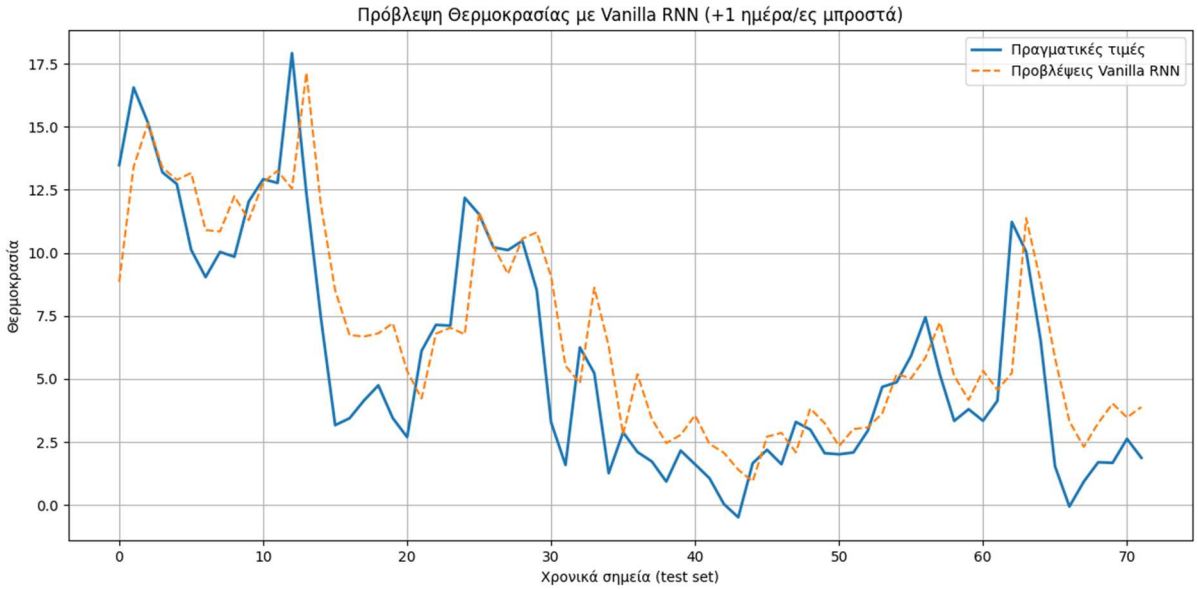
Νευρωνικά δίκτυα

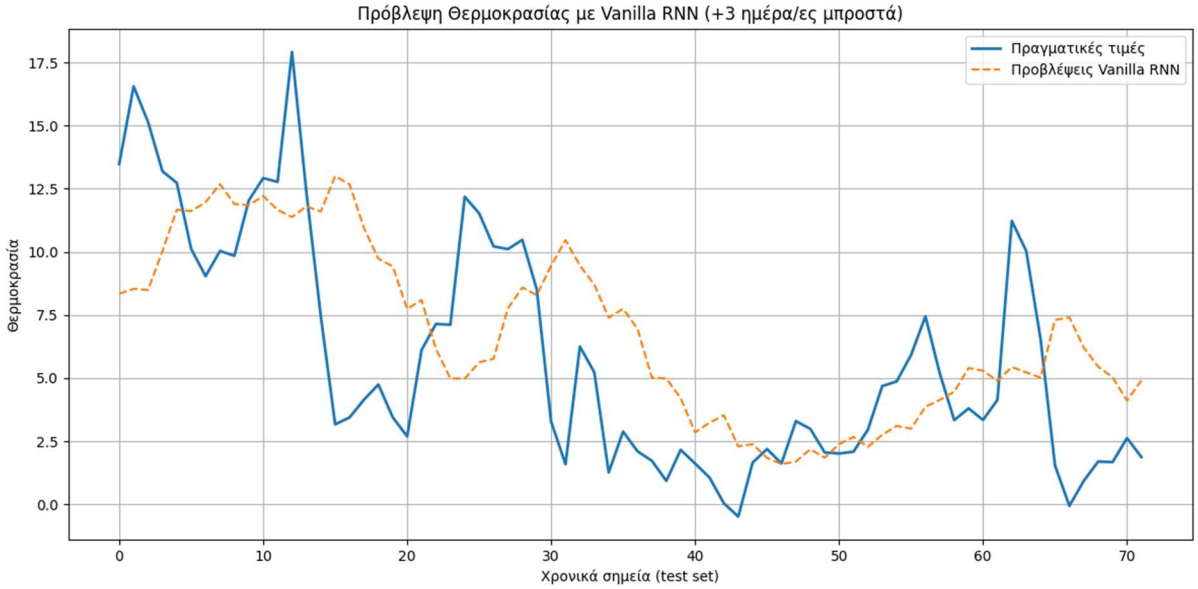
Στο πλαίσιο της παρούσας εργασίας εφαρμόστηκαν τρεις διαφορετικές αρχιτεκτονικές βαθιάς μάθησης: ένα απλό επαναλαμβανόμενο νευρωνικό δίκτυο (Vanilla RNN), ένα RNN με μηχανισμό attention και ένα δίκτυο τύπου Transformer. Και τα τρία μοντέλα εκπαιδεύτηκαν με δεδομένα από τις επτά προηγούμενες ημέρες για την πρόβλεψη της θερμοκρασίας στις επόμενες +1, +3 και +7 ημέρες.

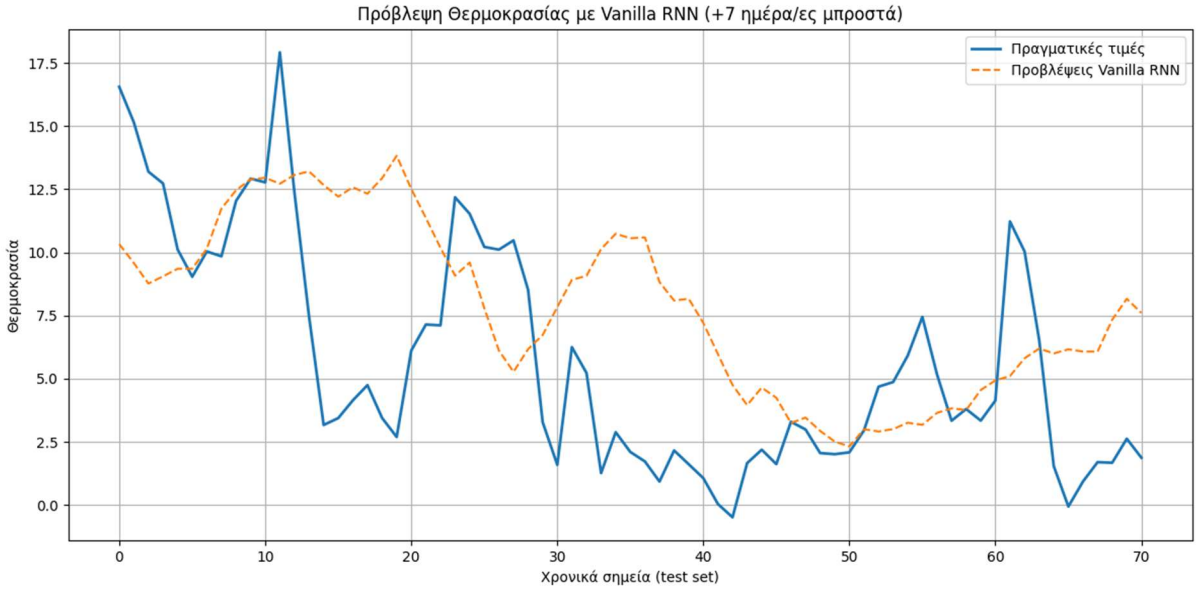
Το Vanilla RNN παρουσίασε πολύ καλή απόδοση στην πρόβλεψη της επόμενης ημέρας (MAE=1.94, RMSE=2.53), όμως όσο μεγάλωνε ο χρονικός ορίζοντας, το σφάλμα αυξανόταν. Το RNN με μηχανισμό attention προσέφερε έναν πιο εξελιγμένο τρόπο αξιοποίησης της πληροφορίας, δίνοντας έμφαση σε συγκεκριμένα χρονικά βήματα της εισόδου κατά την πρόβλεψη. Αυτό είχε ως αποτέλεσμα μεγαλύτερη σταθερότητα στις μεσοπρόθεσμες και μακροπρόθεσμες προβλέψεις (π.χ. στο +7 ημέρες SMAPE 68.23%, ελαφρώς καλύτερο από το Vanilla RNN). Τέλος, το μοντέλο Transformer αξιοποιεί μηχανισμούς attention και positional encoding για να καταγράψει τόσο τοπικές όσο και μακρινές χρονικές εξαρτήσεις. Αν και θεωρητικά υπερέχει, στην πράξη εμφάνισε σημαντικό πρόβλημα στην πρόβλεψη της επόμενης ημέρας (SMAPE=138.16%), ενδεχομένως λόγω υπερβολικής εξομάλυνσης και περιορισμένου μεγέθους δείγματος. Παρ' όλα αυτά, στο χρονικό ορίζοντα +3 ημερών παρουσίασε συγκρίσιμη απόδοση με τα υπόλοιπα μοντέλα, γεγονός που υποδηλώνει τις δυνατότητές του με κατάλληλη βελτιστοποίηση.

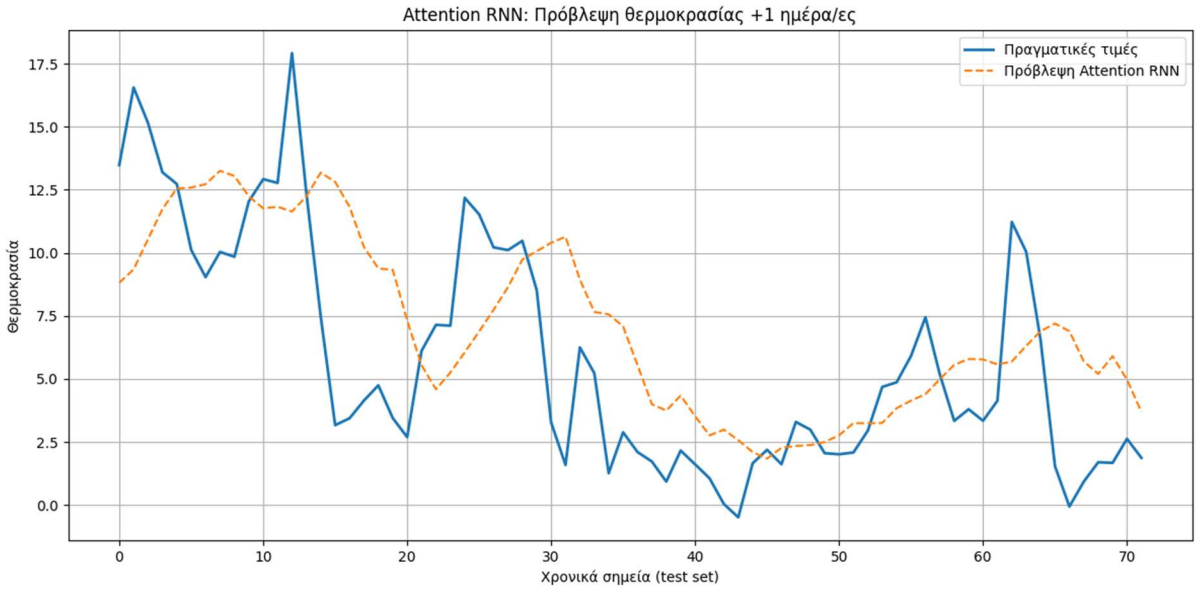
	Model	***MAE***	***RMSE***	***SMAPE***
+1 day	Vanilla RNN	1.94	2.53	45.55%
	Attention RNN	3.04	3.82	57.91%
	Transformer	4.46	5.03	138.16%
+3 days	Vanilla RNN	3.28	4.09	61.43%
	Attention RNN	3.26	4.03	62.37%
	Transformer	3.93	4.64	69.16%
+7 days	Vanilla RNN	3.98	4.94	68.43%
	Attention RNN	3.84	4.67	68.23%
	Transformer	6.27	7.15	87.61%

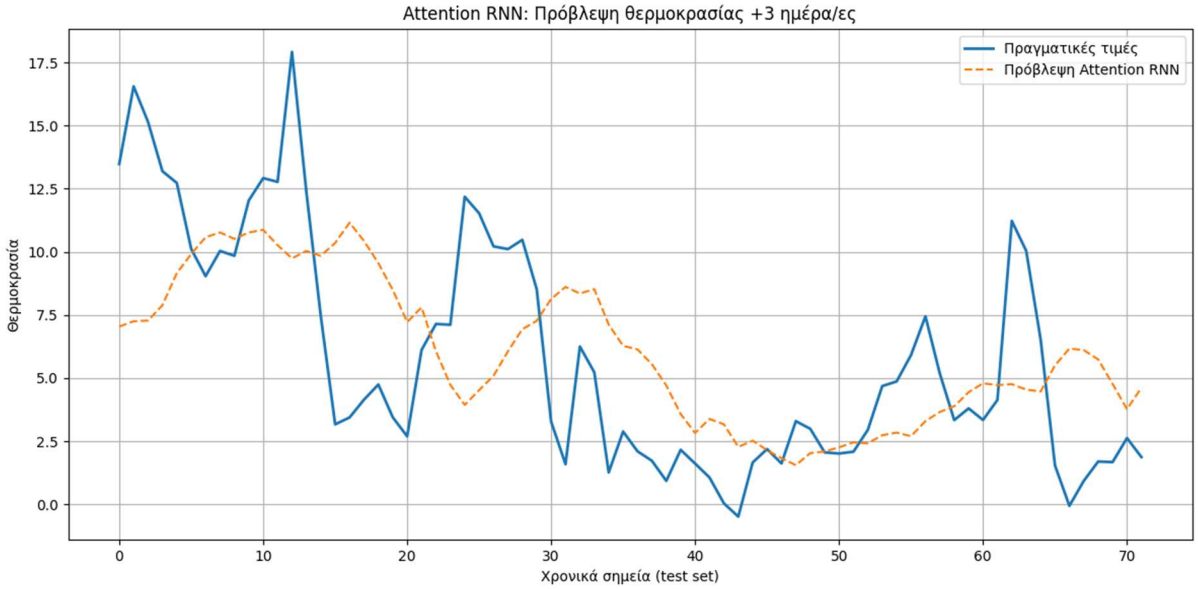
Παρακάτω παρατέθονται διαγράμματα πρόβλεψης των μοντέλων στους ορίζοντες που δόθηκαν (+1 μέρα, +3 μέρες, +7 μέρες)

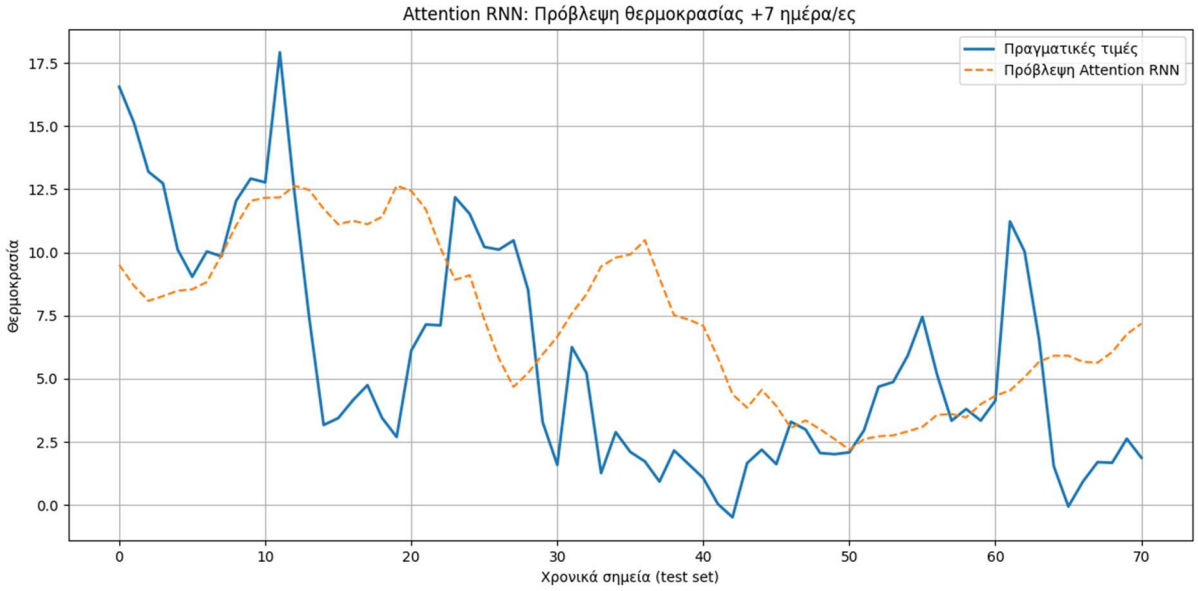

Πρόβλεψη Θερμοκρασίας με Vanilla RNN (+1 ημέρα/ες μπροστά)
Πραγματικές τιμές
Προβλέψεις Vanilla RNN
Θερμοκρασία
Χρονικά σημεία (test set)

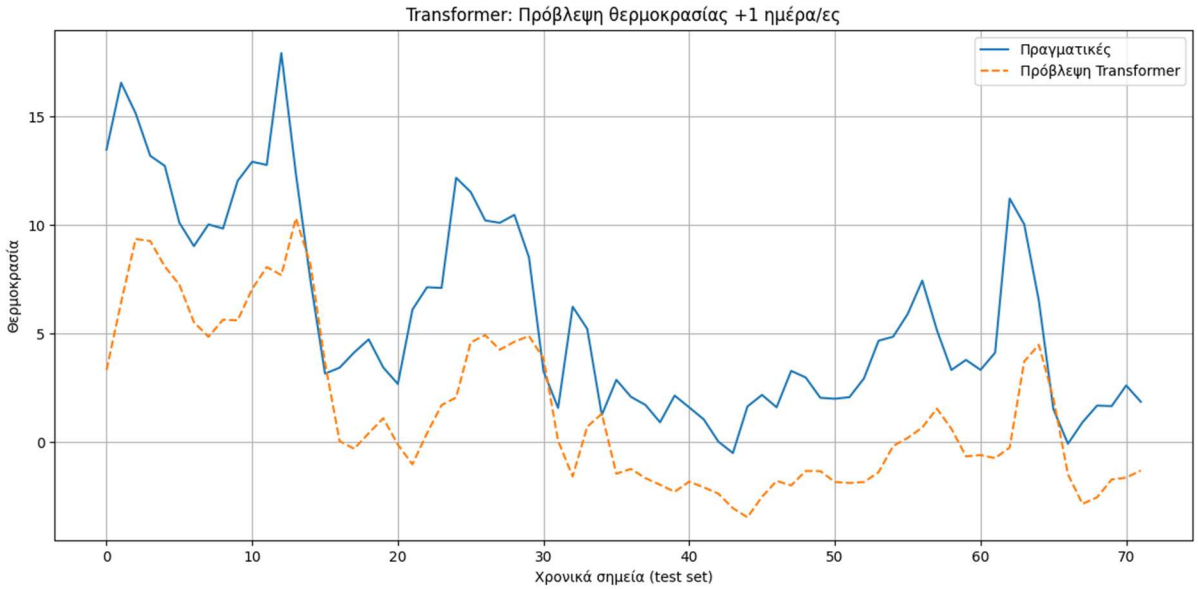

Πρόβλεψη Θερμοκρασίας με Vanilla RNN (+3 ημέρα/ες μπροστά)
Πραγματικές τιμές
Προβλέψεις Vanilla RNN
Θερμοκρασία
Χρονικά σημεία (test set)

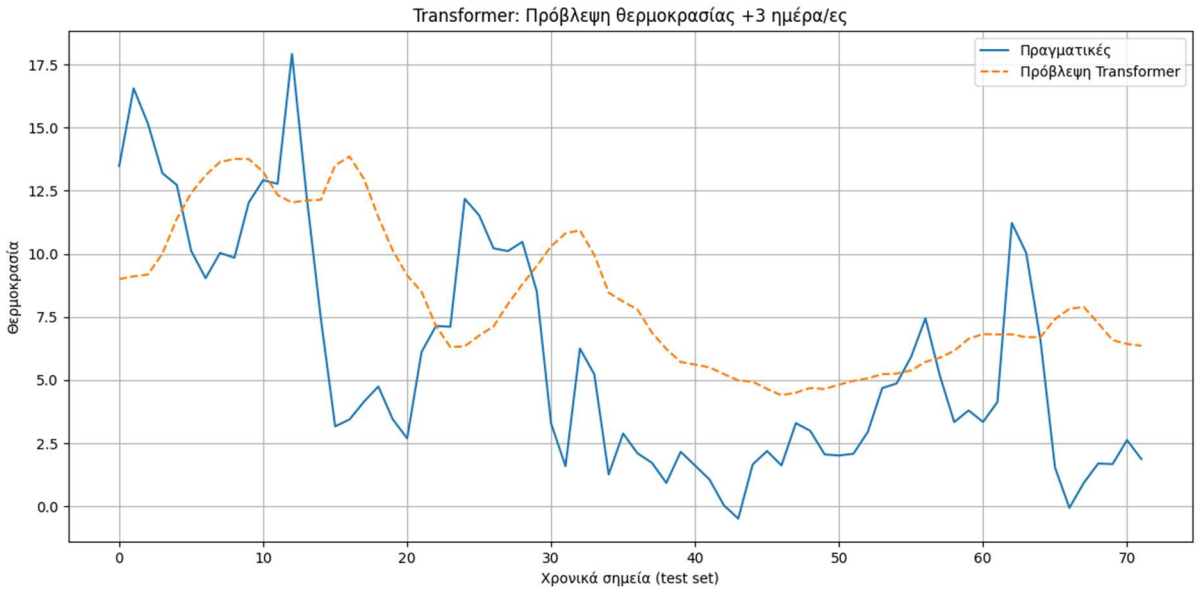

Πρόβλεψη Θερμοκρασίας με Vanilla RNN (+7 ημέρα/ες μπροστά)
Πραγματικές τιμές
Προβλέψεις Vanilla RNN
Θερμοκρασία
Χρονικά σημεία (test set)

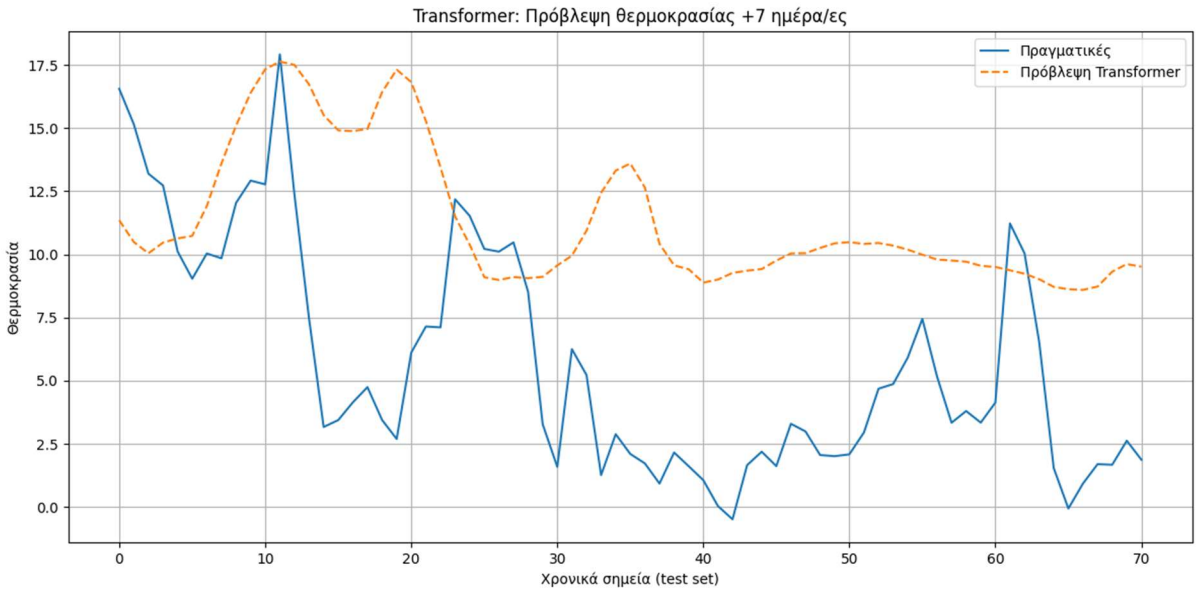

Attention RNN: Πρόβλεψη θερμοκρασίας +1 ημέρα/ες
Πραγματικές τιμές
Πρόβλεψη Attention RNN
Θερμοκρασία
Χρονικά σημεία (test set)


Attention RNN: Πρόβλεψη θερμοκρασίας +3 ημέρα/ες
Πραγματικές τιμές
Πρόβλεψη Attention RNN
Θερμοκρασία
Χρονικά σημεία (test set)


Attention RNN: Πρόβλεψη θερμοκρασίας +7 ημέρα/ες
Πραγματικές τιμές
Πρόβλεψη Attention RNN
Θερμοκρασία
Χρονικά σημεία (test set)


Transformer: Πρόβλεψη θερμοκρασίας +1 ημέρα/ες
Πραγματικές
Πρόβλεψη Transformer
Θερμοκρασία
Χρονικά σημεία (test set)


Transformer: Πρόβλεψη θερμοκρασίας +3 ημέρα/ες
Πραγματικές
Πρόβλεψη Transformer
Θερμοκρασία
Χρονικά σημεία (test set)


Transformer: Πρόβλεψη θερμοκρασίας +7 ημέρα/ες
Πραγματικές
Πρόβλεψη Transformer
Θερμοκρασία
Χρονικά σημεία (test set)