ARISTOTLE UNIVERSITY OF THESSALONIKI
FACULTY OF SCIENCES
SCHOOL OF INFORMATICS
MSC IN DATA AND WEB SCIENCE

Solar Generation Forecasting with Transfer Learning

Πρόβλεψη Παραγωγής Ηλιακής Ενέργειας με Μεταφορά Μάθησης

Dimitrios Mylonas

AEM: 100

Supervisor:

Prof. Ioannis Vlahavas

February 2023

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στο πλαίσιο αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Περίληψη

Η ενέργεια αποτελεί αναπόσπαστο κομμάτι της ανθρώπινης κοινωνίας εδώ και πολλούς αιώνες και η αξιοποίηση της έπαιξε καθοριστικό ρόλο στην επιβίωση και την εξέλιξη του ανθρώπινου είδους. Με το πέρασμα των χρόνων έγινε αντιληπτό ότι οι ανάγκες σε ενέργεια αυξάνονταν και θα συνεχίσουν να αυξάνονται με ραγδαίους ρυθμούς, ενώ οι πόροι που χρησιμοποιούνταν μέχρι πρότινος για την παραγωγή της κάποια στιγμή θα εξαντληθούν. Παράλληλα, τα ορυκτά καύσιμα που αποτελούν τη νούμερο ένα πηγή εκμετάλλευσης για παραγωγή ενέργειας, ο τρόπος εξαγωγής και χρησιμοποίησης τους και η πλήρης εξάρτηση από αυτά για δεκαετίες, έχουν επιβαρύνει ανεπανόρθωτα το περιβάλλον, οδηγώντας στην κλιματική αλλαγή και την καταστροφή αμέτρητων βιοτόπων. Η κλιματική αλλαγή και η περιβαλλοντική επιβάρυνση, οδήγησαν την Ευρωπαϊκή και παγκόσμια κοινότητα στην σύναψη μιας σειράς συνθηκών και συμφωνιών στα τέλη του 20^{ου} αιώνα, με στόχο την επιβολή συγκεκριμένων κατευθύνσεων και οδηγιών στα κράτη μέλη που θα οδηγήσουν σε μείωση και περιορισμό των ρύπων και κατά συνέπεια σε περιορισμό της επιδείνωσης της κλιματικής αλλαγής και της περιβαλλοντικής επιβάρυνσης που προκαλείται από την παραγωγή ενέργειας.

Τα κράτη που δεσμεύθηκαν στις συγκεκριμένες συμφωνίες, θα πρέπει να στραφούν σε εναλλακτικές μορφές ενέργειας, τις οποίες θα προσθέσουν στην ενεργειακή τους αγορά και στο ενεργειακό τους μείγμα με στόχο να μειωθούν τα ποσοστά εκπομπών διοξειδίου του άνθρακα στην ατμόσφαιρα, πράγμα που οφείλεται στην καύση των ορυκτών καυσίμων. Η Σύμβαση-Πλαίσιο των Ηνωμένων Εθνών για την Κλιματική Αλλαγή, το Πρωτόκολλο του Κιότο και η Ευρωπαϊκή Πράσινη Συμφωνία είναι μερικές από τις πιο σημαντικές συνθήκες που συμφωνήθηκαν στις αρχές της δεκαετίας του 90’ και έπειτα, με κάποιες από αυτές να βρίσκονται ακόμα σε ισχύ. Η Ελλάδα, όπως και οι υπόλοιπες Ευρωπαϊκές χώρες σήμερα, είναι υποχρεωμένη να προσαρμόσει την ενεργειακή της αγορά με βάση το Ευρωπαϊκό Μοντέλο Στόχος.

Ηλιακή, αιολική, γεωθερμική, βιομάζα, ενέργεια από τη θάλασσα και υδροηλεκτρική ενέργεια, είναι σήμερα οι έξι βασικές ανανεώσιμες πηγές ενέργειας. Ήπιες, πράσινες ή ανανεώσιμες πηγές ενέργειας (ΑΠΕ), είναι οι πηγές εκείνες που προέρχονται από το περιβάλλον και από τα διάφορα φυσικά φαινόμενα και μπορούν να εκμεταλλευθούν για την παραγωγή ενέργειας. Για την εκμετάλλευσή των ήπιων πηγών ενέργειας, δεν απαιτείται καύση, εξόρυξη ή οποιαδήποτε άλλη ενεργητική παρέμβαση, ενώ παράλληλα αποτελούν «καθαρές» μορφές ενέργειας που δεν επιβαρύνουν το περιβάλλον και δεν εξαντλούνται.

Ωστόσο, οι πράσινες πηγές ενέργειας δεν έχουν την ίδια αποδοτικότητα με τις παραδοσιακές μορφές και για το λόγο αυτόν πρέπει να αξιοποιούνται σε πολύ μεγάλες ποσότητες, πράγμα που απαιτεί την εκμετάλλευση τεράστιων εκτάσεων για την κατασκευή των μονάδων παραγωγής τους. Αποτέλεσμα αυτού είναι η καταστροφή δασών ή άλλων βιοτόπων και τα ιδιαίτερα αυξημένα έξοδα για τις εταιρείες παραγωγής ενέργειας, συγκριτικά με τα έξοδα κατασκευής μιας μονάδας παραγωγής ενέργειας από ορυκτά καύσιμα. Επιπλέον, η αποδοτικότητα των ήπιων μορφών ενέργειας είναι άρρηκτα συνδεδεμένη με τις καιρικές

συνθήκες, συνεπώς μπορεί να καταστεί απρόβλεπτη και ασταθής, ενώ οι περιβαλλοντικές συνθήκες ορισμένων χωρών δεν επιτρέπουν την αξιοποίηση όλων των ανανεώσιμων μορφών ενέργειας. Για το λόγο αυτό, απαιτούνται μονάδες αποθήκευσης ενέργειας υψηλής απόδοσης, οι οποίες αυξάνουν εκ' νέου το κόστος του εγχειρήματος, ωστόσο προστατεύουν την ενεργειακή αγορά από ξαφνικές αποκλίσεις μεταξύ ζήτησης και παραγωγής.

Γίνεται αντιληπτό, ότι η επικράτηση των ήπιων μορφών ενέργειας και η πλήρης αντικατάσταση των συμβατικών μορφών από αυτές είναι εξαιρετικά δύσκολη, απαιτεί την δαπάνη τεράστιων χρηματικών ποσών και κυριότερα την εύρεση λύσεων για την μεγιστοποίηση της αποδοτικότητας τους και την εξάλειψη των όποιων κινδύνων και ρίσκων ελλοχεύει η πιθανή εξάρτηση της αγοράς ενέργειας σε αυτές. Παρ' όλες τις δυσκολίες, τους περιορισμούς και τα έξοδα που απαιτούνται, η πράσινη μετάβαση είναι πιο αναγκαία από ποτέ και για το λόγο αυτό, η παγκόσμια επιστημονική κοινότητα έχει στραφεί στην εύρεση λύσεων και βέλτιστων πρακτικών για την καλύτερη αξιοποίηση των πράσινων μονάδων παραγωγής ενέργειας. Η δημιουργία σύνθετων συστημάτων πρόβλεψης της παραγωγικότητας των μονάδων, αποτελεί ένα από τα βασικά σημεία ενδιαφέροντος στο πεδίο της έρευνας, καθώς πρόκειται για ένα πολυπαραμετρικό και πολυσύνθετο πρόβλημα το οποίο όμως μπορεί να δώσει οριστικές και καθοριστικές λύσεις στις προσπάθειες πράσινης μετάβασης. Τέτοια εργαλεία μπορούν να παίξουν καθοριστικό ρόλο στην οικονομική βιωσιμότητα των μονάδων παραγωγής ενέργειας, καθώς επίσης και στη διατήρηση ισορροπίας μεταξύ προσφοράς και ζήτησης στην αγορά.

Η μηχανική μάθηση, ως πεδίο της Τεχνητής Νοημοσύνης, βασίζεται στην ιδέα ότι τα συστήματα μπορούν να μάθουν από τα δεδομένα, να εντοπίσουν μοτίβα και να πάρουν αποφάσεις με βάση αυτά, καθώς και να βελτιωθούν με την εμπειρία. Η μορφή, η ποιότητα και η ποσότητα των δεδομένων είναι καθοριστικοί παράγοντες για την επιλογή του αλγορίθμου μηχανικής μάθησης που ταιριάζει στο εκάστοτε πρόβλημα, καθώς και για την αποδοτικότητα αυτού. Οι αλγόριθμοι μηχανικής μάθησης ενδείκνυνται για την εκμετάλλευση μεγάλου όγκου δεδομένων και την αξιοποίηση εξωγενών παραμέτρων και μεταβλητών για την επίλυση προβλημάτων και ο τομέας της ενέργειας αποτελεί ένα τέτοιο πρόβλημα. Εκατοντάδες διαφορετικές προσεγγίσεις, αλγόριθμοι και τεχνικές μηχανικής μάθησης έχουν δοκιμαστεί και συνεχίζουν να μελετώνται με στόχο την εύρεση πρακτικών για βέλτιστες προβλέψεις σχετικά με την παραγωγή ή την κατανάλωση ενέργειας, ή ακόμη και την ορθή λειτουργικότητα και συντήρηση των μονάδων παραγωγής.

Η παρούσα διπλωματική, μελετάει την δημιουργία προεκπαιδευμένων μοντέλων μηχανικής μάθησης και πιο συγκεκριμένα νευρωνικών δικτύων και την ικανότητα αυτών να κάνουν πιο ακριβείς προβλέψεις της παραγόμενης ηλιακής ενέργειας, συγκριτικά με μοντέλα που έχουν εκπαιδευτεί σε λιγότερα δεδομένα. Πιο αναλυτικά, γίνεται χρήση δύο συνόλων δεδομένων μεγάλου όγκου, στη μορφή χρονοσειρών, με ωριαίες καταγραφές παραγωγής ηλιακής ενέργειας για πάνω από 5 χρόνια, για ένα σύνολο Ευρωπαϊκών χωρών. Τα μοντέλα εκπαιδεύονται στο σύνολο των χωρών και στη συνέχεια καλούνται να κάνουν προβλέψεις για τις χώρες αυτές, καθώς και για μεμονωμένες χώρες που έχουν λίγα ή καθόλου δεδομένα στη διάθεση τους για εκπαίδευση. Για την αξιολόγηση των μοντέλων χρησιμοποιούνται οι μετρικές Mean Absolute Error και Root Mean Squared Error, ενώ γίνεται εκτενής σύγκριση με άλλα μοντέλα που χρησιμοποιήθηκαν σαν baselines και τα οποία έχουν εκπαιδευτεί σε κάθε χώρα ξεχωριστά.

Πιο συγκεκριμένα, μετά από εκτενή προ επεξεργασία των δεδομένων για την αφαίρεση των missing values, την κανονικοποίηση των τιμών και την επιλογή των χωρών με την μεγαλύτερη ποικιλομορφία δεδομένων, τα δύο σύνολα δεδομένων συνενώθηκαν, οδηγώντας στο τελικό dataset το οποίο αποτελείται από περίπου 350.000 παρατηρήσεις και τέσσερα διαφορετικά χαρακτηριστικά. Η θερμοκρασία, ο μήνας και η ηλιακή ακτινοβολία είναι τα τρία βασικά χαρακτηριστικά που αξιολογούνται από το μοντέλο για να προβλέψει την μεταβλητή στόχο που είναι η παραγωγή ηλιακής ενέργειας. Το σχηματισμό του συνόλου δεδομένων ακολούθησε η εφαρμογή μεθόδου windowing, μια τεχνική που χρησιμοποιείται εκτενώς σε προβλήματα με δεδομένα χρονοσειρών και τα μετατρέπει σε μορφή επεξεργάσιμη από ένα νευρωνικό δίκτυο. Στα πλαίσια της εργασίας αποφασίστηκε οι προβλέψεις των μοντέλων να αφορούν τις επόμενες 24 ώρες.

Τα πειράματα ξεκίνησαν με ένα πολύ απλό Convolutional νευρωνικό δίκτυο στο οποίο έγινε fine tuning και καταγράφηκαν τα πρώτα αποτελέσματα. Στη συνέχεια, ακολούθησαν πολλαπλές δοκιμές και προσθήκες, όσον αφορά το βάθος του νευρωνικού και το είδος των layers που ενδέχεται να προστεθούν. Έγιναν δοκιμές για τον ιδανικό αριθμό convolutional layers που θα απαρτίζουν το μοντέλο, αντίστοιχες δοκιμές για dense layers, ενώ μεταξύ πολλών layers προστέθηκαν Dropout και BatchNormalization layers. Η προσθήκη νέων layers και η επιλογή των τιμών των υπερπαραμέτρων έγινε με γνώμονα την όσο το δυνατόν μεγαλύτερη πτώση της τιμής του Root Mean Squared Error (RMSE). Επιπλέον, προστέθηκαν residual layers όπου κρίθηκε σκόπιμο, ενώ εξετάστηκε και η τιμή του learning rate, σε μια προσπάθεια βελτίωσης του χρόνου εκπαίδευσης. Έπειτα από εκατοντάδες πειράματα και δοκιμές, προέκυψαν τρεις βαθιές αρχιτεκτονικές νευρωνικών δικτύων, οι οποίες φαίνεται να αποδίδουν αποτελεσματικά.

Ως baselines, επιλέχθηκαν ένα απλό Convolutional Neural Network (simple CNN), δύο πολυχρησιμοποιημένα στατιστικά μοντέλα, ο Exponential Smoothing και η Theta Method, καθώς και δυο naive προσεγγίσεις. Για τα δύο στατιστικά μοντέλα, εφαρμόστηκε fine tuning για την καταλληλότερη επιλογή των υπερπαραμέτρων τους και στη συνέχεια εκπαιδεύτηκαν – όπως και το simple CNN σε κάθε μία από τις επιλεγμένες χώρες ξεχωριστά. Ο Exponential Smoothing και το Simple CNN ξεχώρισαν μεταξύ των baselines και για το λόγο αυτό επιλέχθηκαν για περαιτέρω πειράματα και λεπτομερή σύγκριση με τα τρία προεκπαιδευμένα νευρωνικά δίκτυα. Τα προεκπαιδευμένα νευρωνικά μοντέλα απέδωσαν καλύτερα σε μέσο όρο, καθώς και ατομικά σε κάθε μία από τις χώρες στις οποίες είχαν εκπαιδευτεί, λαμβάνοντας υπόψη πέρα από την RMSE και τις τιμές της Mean Absolute Error (MAE). Το καλύτερο μοντέλο μεταξύ των τριών όσον αφορά τις τιμές της RMSE, πέτυχε μέσο όρο σφάλματος κατά 2.5 φορές μικρότερο από το μέσο όρο του Exponential Smoothing και του simple CNN, ενώ το πιο αποδοτικό μοντέλο με βάση τις τιμές της MAE πετυχαίνει μέσο όρο πάνω από 3 φορές μικρότερο των baselines.

Τα αποτελέσματα έδειξαν πως η εκπαίδευση ενός νευρωνικού δικτύου σε ένα μεγάλο όγκο δεδομένων με μεγαλύτερη ποικιλία και συνδυασμού τιμών, μπορεί να οδηγήσει σε γενίκευση και κατ' επέκταση σε μεταφορά γνώσης και δυνατότητα πιο αποτελεσματικών προβλέψεων για τις χώρες στις οποίες έχει εκπαιδευτεί. Τα προεκπαιδευμένα μοντέλα, όταν κλήθηκαν να κάνουν προβλέψεις για τις χώρες οι οποίες ανήκουν στο σύνολο εκπαίδευσή τους, ήταν πιο αποτελεσματικά από τα baselines, καθώς και από νευρωνικά με τις ίδιες ακριβώς αρχιτεκτονικές, αλλά εκπαιδευμένα ατομικά σε κάθε χώρα. Όσον αφορά τις προβλέψεις σε τελείως άγνωστες χώρες, τα αποτελέσματα ήταν πιο ισορροπημένα. Τα προεκπαιδευμένα μοντέλα ξεπέρασαν τις επιδόσεις των baselines σε 2 χώρες, ενώ υστερούσαν στις άλλες 2. Συνολικά, ο μέσος όρος σφαλμάτων κάθε προ εκπαιδευμένου μοντέλου για τις προβλέψεις στις άγνωστες χώρες ήταν καλύτερος των baselines με βάση την MAE και χειρότερος από τους μέσους όρους του Exponential Smoothing με βάση την RMSE. Με άλλα λόγια, τα προ εκπαιδευμένα μοντέλα ξεπέρασαν σε μέσο όρο το simple CNN στις άγνωστες χώρες, ενώ υπήρξε ισορροπία στη σύγκριση τους με τον Exponential Smoothing. Συμπερασματικά, οι προβλέψεις στις άγνωστες χώρες, φαίνεται να είναι ισορροπημένες και άρρηκτα συνδεδεμένες με τα δεδομένα κάθε χώρας. Για το λόγο αυτό, απαιτείται περισσότερη διερεύνηση και στατιστικά τεστ προκειμένου να καταλήξουμε σε οποιοδήποτε συμπέρασμα.

Τα περιεχόμενα της εργασίας οργανώνονται σε 5 κεφάλαια με το πρώτο να αποτελεί την σύντομη εισαγωγή στο θέμα και τελευταίο τα αντίστοιχα συμπεράσματα. Στο κεφάλαιο 2 γίνεται ανάλυση του τομέα της Μηχανικής μάθησης, δίνονται ορισμοί των βασικών προβλημάτων και παρουσιάζονται κάποιοι από τους βασικότερους αλγορίθμους όπως τα Decision Trees, Linear Regression και Anomaly Detection. Επιπλέον, αναλύονται τα νευρωνικά δίκτυα, καθώς και το πεδίο της πρόβλεψης χρονοσειρών που είναι τα δύο βασικά κομμάτια ενασχόλησης της παρούσας διπλωματικής. Στη συνέχεια, στο κεφάλαιο 3, παρουσιάζεται και αναλύεται το πεδίο εφαρμογής της εργασίας, δηλαδή ο τομέας της ενέργειας. Εξετάζονται λεπτομερώς οι ανανεώσιμες πηγές ενέργειας και η σημαντικότητα αυτών, ενώ παρουσιάζονται και οι κινητοποιήσεις της Ευρωπαϊκής και παγκόσμιας κοινότητας τις τελευταίες δεκαετίες, στα πλαίσια της προσπάθειας για πράσινη μετάβαση. Τέλος, στο κεφάλαιο 4 αναλύεται εκτενώς η πειραματική διαδικασία της εργασίας. Γίνεται παρουσίαση και ανάλυση των δεδομένων και της επεξεργασίας που υπέστησαν, λεπτομερής ανάλυση του τρόπου ανάπτυξης των τριών αρχιτεκτονικών καθώς και παράθεση των αποτελεσμάτων και ευρημάτων.

Abstract

The transition to alternative forms of energy becomes more and more important through the years due to resources depletion and mainly due to climate change. For this reason, the European Union forced its members to comply with a strict target model which is targeting to transform Europe's economy to be resource efficient, environmentally sustainable and competitive. Wind, hydropower, geothermal, biomass, marine and solar energy constitutes the six main renewable energy sources and their exploitation differs from country to country as their efficiency are fully correlated with the geographical and climate conditions. Thus, for each European country specific energy targets have been set, concerning their respective climate conditions and a set of other parameters.

Solar energy is a very promising energy source and its use has been growing rapidly in the recent years, especially in countries like Greece with intense sunshine and high temperatures during most months of the year. However, the installation of solar energy units remains a quite expensive procedure and their efficiency can be unstable and unpredictable as it is highly related with weather conditions. Solar radiation and temperature are two extremely important weather features which affect directly the performance of a solar panel and are difficult to predict accurately. Thus, finding efficient ways to predict weather conditions or better to forecast solar units' productivity in correlation with the weather conditions can be revolutionary. During the last decades, the scientific community has turned its focus on finding ways to maximize and optimize the performance of renewable energy units and the forecasting of their production has become of the main topics. Efficient forecasting can lead to the coveted balance between the supply and demand in the energy market, which is one of the most critical points in the global effort for the transition from the conventional to alternative energy sources.

This dissertation focuses on the development of pretrained Convolutional neural networks which can efficiently forecast solar energy production. More specifically, an investigation takes place concerning the ability of these pretrained models to be more efficient than models which are trained with small volumes of data. In other words, neural networks are trained on a large timeseries dataset with solar and weather data from several European countries and then make predictions firstly for the dataset's countries and then for countries with a few data. The experiments resulted in three deep neural network architectures which are mainly composed of Convolutional and Dense layers. At the same time, four different statistical models and one simple CNN were trained on each country's data separately and were used as baselines, with Exponential Smoothing and simple CNN being the most efficient one. The results showed that indeed the neural networks when trained on a massive and qualitative dataset with a variety of different examples from several countries can generalize and transfer their knowledge to make more accurate predictions for these countries or even for new countries with no data or a few data available. All of the three pretrained neural networks outperform by far the baselines when they did predictions for the countries on which they got trained, with the error being reduced by up to 3 times in some cases. Also, they performed better than neural networks with exactly the same architectures and which were trained

individually in each country. When it comes to predictions for unknown countries, the results were interesting but more balanced and further investigation and statistical tests are required in order to draw more certain conclusions.

Acknowledgements

I am immensely grateful to Prof. Ioannis Vlachavas, my dissertation supervisor, for his assistance and guidance. Also, I would like to express my deep appreciation and gratitude to co-supervisor and PhD candidate Ioannis Pierros for his invaluable support and patience throughout my research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The energy demand worldwide is increasing steadily and over the past twenty years, it seems that the global energy consumption has almost doubled [1]. Despite this, most of the countries continue to rely on conventional energy sources, which exacerbate the greenhouse effect and the climate change, while there is always the risk of exhausting these resources. These are some of the main reasons which led the global community to turn its focus on alternative energy sources and on finding ways to optimize their exploitation.

As renewable energy source can be considered any kind of energy source which is based on the environment's natural phenomena and is both sustainable and unlimited. Solar, wind, biomass, hydropower, geothermal and marine energy are the main six renewable energy forms and they are produced either from natural phenomena, or directly or indirectly from the sun. The exploitation and the adaption of alternative energy forms like the aforementioned is more urgent than ever and for this reason the European Union established the European Target Model aiming to control the functioning of the energy market within the Member states.

However, renewable energy exploitation has several disadvantages. The solar units that must be installed have significantly high equipment and maintenance costs and also require nature intervention as they should be installed in large areas with specific natural characteristics (forests, sea, rivers etc.). Except of the installation difficulties and the environmental issues which make the total transition to alternative energy controversial, the unpredictable nature of renewable energy sources results to unstable energy production. This probably constitutes the main disadvantage of renewable energy sources compared to conventional ones, as the need for balance between supply and demand in energy market is something non-negotiable and only with stable and predictable energy production can be maintained. The unstable nature of renewable energy sources results from their high correlation with the weather conditions, such as wind, sunshine intensity, temperature etc. which are difficult to predict accurately.

It is clearly seen that the prevalence of renewable energy sources over the conventional ones is extremely difficult and requires the elimination of the aforementioned risks and problems. In the last decades, research efforts have shifted towards enhancing the utilization of renewable energy with energy production forecasting being one of the most studied topics. Several statistical and machine learning algorithms and techniques have been tested focusing on accurate forecasting concerning the production of renewable energy units, while the problem can be even more complicated if someone wants to predict the productivity of units with no or little history data.

The main focus of this thesis, is to develop one or more pretrained machine learning models which can generalize and transfer their knowledge and make efficient predictions for the countries on which they were trained and also to unseen data. These models are trained on a large timeseries solar generation dataset from several European countries and make predictions for these countries and for new countries with small volumes of available data or no data at all. Additional features concerning the weather conditions such as temperature, month of the year and solar radiation were also used to feed the model with more details. Furthermore, the three neural networks that came up after the testing and fine-tuning procedures are compared with baseline models which are trained in individual countries.

The structure of this thesis is the following. An extensive analysis of machine learning is given in Chapter 2 with details about algorithms, metrics, techniques and a focus on neural networks and time series forecasting which are the main subjects of this dissertation. In Chapter 3 the renewable energy sources are presented in detail with references to their advantages and disadvantages. Furthermore, the global mobilization concerning the climate change is described and the actions taken by the global community are analyzed. Also, an analysis of the Greek energy market is given and the importance of forecasting the energy production is mentioned too. Subsequently, in Chapter 4, the main work of the current thesis is presented including the dataset preprocessing and the neural networks' construction. Moreover, in the same Chapter, the baselines analysis and the results comparison take place. Finally, Chapter 5 provides a summary of the work done and also some ideas for future research and potential growth on this specific topic.

Chapter 2

Machine Learning

Learning is the process of gaining new knowledge, skills and behaviors and this process starts for human beings since their birth. One can learn by a single event or easily from repeated experiences and patterns by linking each result to the action that preceded it. Except humans, animals and plants, the mastery of learning is something that concerns and can be done by some machines as well.

The science of Artificial Intelligence (AI) makes it possible for some computers to obtain knowledge and learn through experience, adjust to new input data and perform human-like tasks such as image and voice recognition, decision making, content creation etc. In other words, Artificial Intelligence refers to the process that simulates human intelligence by computers which are programmed to act, behave and think like humans. Its fundamental principle is that human reasoning and way of thinking can be easily defined in a way that a computer can imitate and perform simple or more complex tasks. These machines are programmed to take actions which are justified and explainable and mainly to have the best chance of achieving a specific goal through these actions.

Machine Learning is a subfield of Artificial Intelligence which is based on the idea that systems can learn from the data, locate patterns in them and make decisions with little or no human intervention and at the same time to be improved with time through experience. Similar to how a human brain acquires knowledge, machine learning relies on input data. It tries to understand entities, patterns and the correlations between them. One could say that the following aspects, are the main components and key elements of Machine Learning:

- **Data set**: Computers needs a lot of data to learn from and make decision based on them. A data set is an organized collection of related data or records with specific characteristics and types concerning a specific topic. Machine Learning models will be trained and validated with these data so the bigger the data set, the better the models will learn.
- **Models and algorithms**: Machine learning models and algorithms are a mixture of statistics, math and programming. An algorithm can be considered as a mathematical or logical set of rules or a problem-solving operation which use computational methods to extract information from the data. Models are trained over a data set in a way that makes them able to identify certain types of data and make predictions on those or similar data. There is a variety of different types of Machine Learning models and algorithms that can be chosen and used, depending on various aspects such as the nature of data, the available resources, the problem that we are trying to solve etc. There are three main types of Machine Learning which differs on the way that algorithms are being trained and the kind of the data that they ingest: Supervised Learning, Unsupervised Learning and Reinforcement Learning. Additionally, an

extremely important type of Machine Learning models is the Neural Networks which falls under the category of Deep Learning.

- **Evaluation**: Each machine learning model should be able to be improved and for that reason we need to understand its performance and its weaknesses and strengths. Different evaluation measures and metrics have been developed and used depending on the type of Machine Learning problem each time. The process of measuring the performance of a model is known as model's evaluation and it is extremely important for the model's improvement.

In the next subsections, the two main types of Machine Learning will be extensively analyzed. The analysis concerns the basic philosophy of each type, the main algorithms and the way that they work and their evaluation process and metrics. Additionally, considerable attention is given to the field of Deep Learning where the fundamental concepts and some basic types of Neural Networks are explained. Finally, an introduction to Time Series Forecasting is provided. In Figure 1, an overview of machine learning tasks and applications is given.

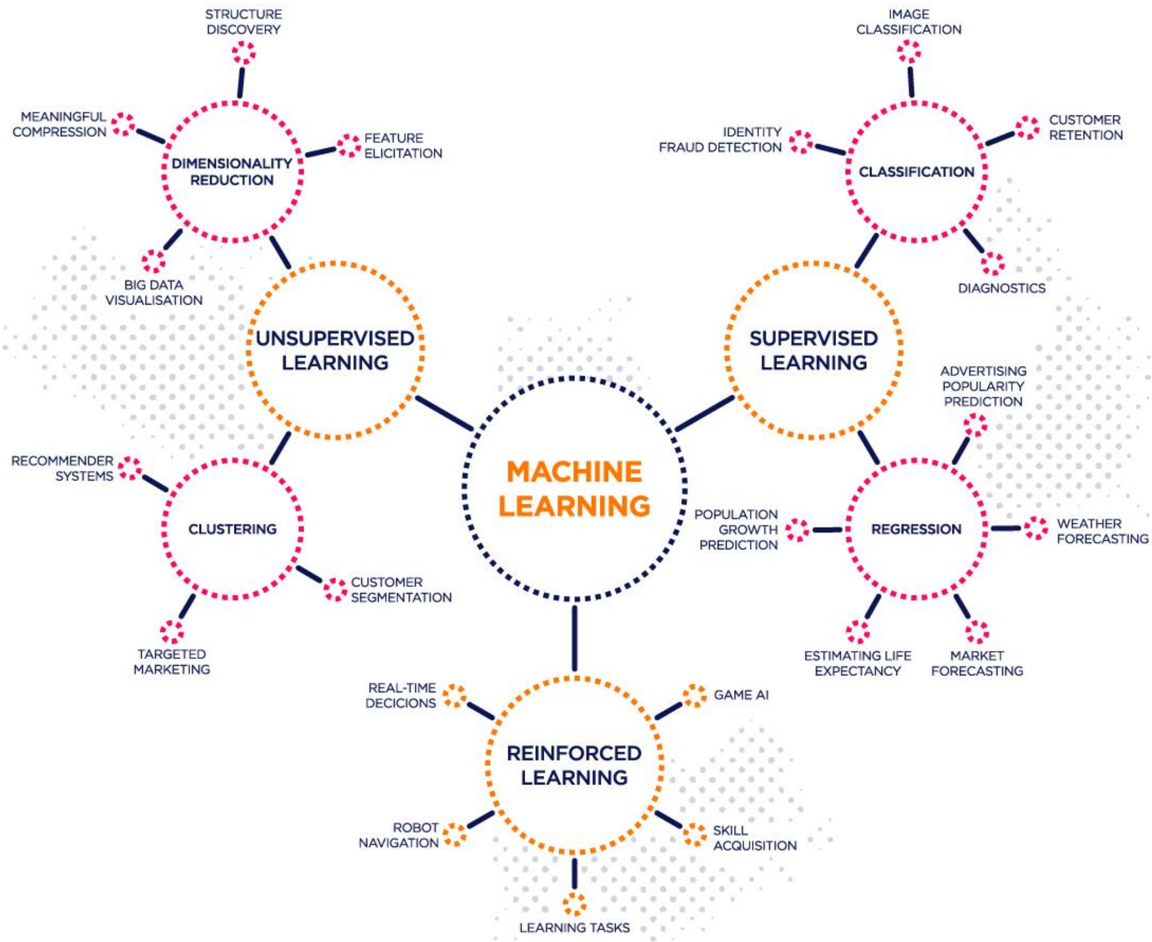

Figure 1: Machine learning tasks and applications

2.1 Supervised Learning

In Supervised Learning algorithms learn from labeled data [2]. In a labeled training data set the corresponding output of each input data is provided. In this type of Machine Learning task, algorithms ingest inputs and the appropriate outputs and, in this way, detect correlations and patterns between them. The goal is to learn a general model (function) that maps outputs to the inputs. After the training process, this model should be able to classify an unseen input to the appropriate output class. In other words, a model (classifier) is used to predict the class that a given input belongs to. In a more formal and mathematical perspective, Supervised Learning can be described as a set of the following ingredients:

- Input variables (X)
- Output variable (Y)
- A target function $Y = f(X|P) + E$, where $f(\cdot)$ is the model, P the parameters and E the error of the model's estimations.
- Model = Algorithm (Data)

During the learning process the algorithm tries to optimize the parameters in order to minimize the error and thus to optimize the model's predictions to be as close as possible to the true values. It should be mentioned that machine learning models and algorithms are not the same concept. The model is the final representation which resulted from the learning process and the specific and defined parameters and data, while the algorithm is the model's building process.

Outputs can be divided in one or more classes which depends on the kind of the data and the subject that is investigated. The models' accuracy is calculated by comparing the algorithm's predictions with the correct outputs and then it can be improved by adjusting the model's parameters. There are two types of Supervised Learning algorithms regarding the desired output:

- **Classification**: the algorithm can predict a discrete value which is in the form of an integer quantity.
- **Regression**: the algorithm can predict a continuous value if it is in the form of a class label probability.

Both types work with labeled data. The difference between them is that they are used for different problems with different kind of outputs. In the next subsections, several Classification and Regression models are presented and analyzed. An example of each of the two Supervised Learning problems can be seen in Figure 2.

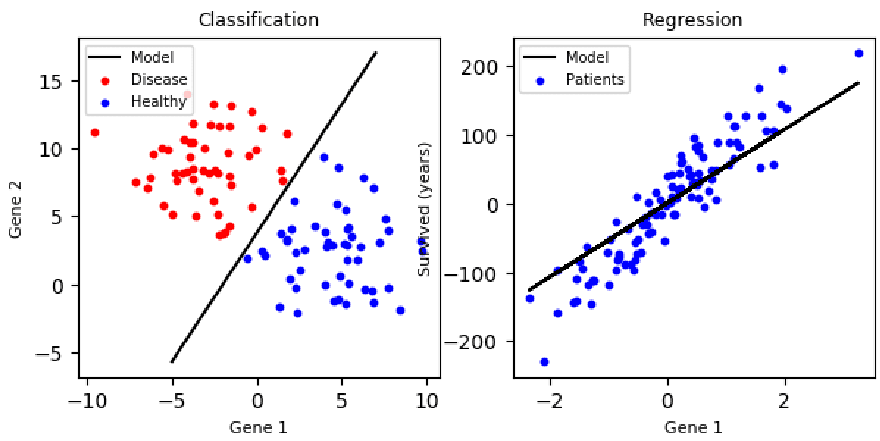

Figure 2: An example of Classification (left) and Regression (right) problems

2.1.1 Evaluation Metrics

All machine learning models need an evaluation process and a metric to measure and judge their accuracy. There are several evaluation metrics for both types of problems. Each evaluation metric measures different qualities and manifests different outcomes. The most common metrics for each type of supervised learning problems are presented below.

<u>Regression</u>

- **Mean Squared Error (MSE)**
  It calculates the average of the squared difference between the target value and the value that the regression model predicted. MSE can be optimized better as it is differentiable, though it is fundamentally more prone to outliers than other metrics due to the squaring factor. It is calculated as below:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(y_j - \check{y}_j)^2$$

- **Mean Absolute Error (MAE)**
  MAE finds the average of the difference between the ground truth and the model's predicted values. In contrast to MSE, MAE is non-differentiable and it is more robust towards outliers since it doesn't exaggerate larger errors. It should be mentioned that MAE doesn't give us an insight of the error's direction. It is represented as below:

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|y_j - \check{y}_j|$$

- **Root Mean Squared Error (RMSE)**
  It finds the square root of the average of the squared difference between the target value and regression model's predicted values. RMSE retains the differentiability as MSE and scale factors are normalized, so in the case of outliers it is less prone to struggle.

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \check{y}_j)^2}$$

Classification

A fundamental tool in the evaluation process of Classification problems is the **Confusion Matrix.** The confusion matrix is not a performance metric, but is the basis on which many models evaluate their results in collaboration with some appropriate metrics. It is nothing but a tabular visualization of the model predictions against the ground truth. Each column of the confusion matrix represents the examples in the actual class, while each row of this matrix shows the instances in the predicted class. In other words, each of the four cells in the confusion matrix constitutes an evaluation factor, as seen in Figure 3. These factors are defined as follows:

- **True Positive (TP):** the number of positive class examples that were predicted correctly by the model.
- **True Negative (TN):** the number of negative class examples that were predicted correctly by the model.
- **False Positive (FP):** the number of positive class examples that were predicted incorrectly by the model.
- **False Negative (FN):** the number of negative class examples that were predicted incorrectly by the model.

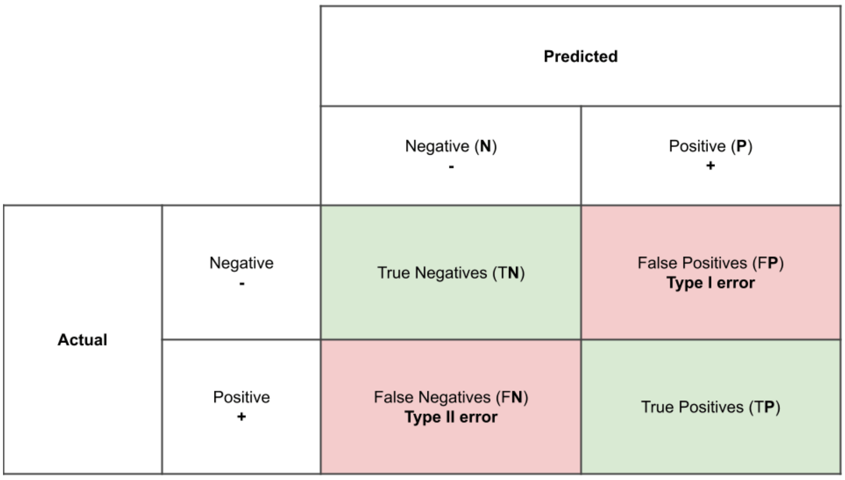

Figure 3: Confusion Matrix for Binary Classification

Accuracy

It constitutes one of the most common and simplest metrics to implement and use in order to evaluate a machine learning model for supervised classification problems. It is defined as the fraction of the model's correct prediction to the total number of model's predictions, multiplied by 100.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP + TN + FP + FN}$$

Precision

This metric can be defined as the fraction of true positives to total positives predicted. In contrast with accuracy, precision is useful for unbalanced datasets. Its measures how many of the correctly predicted examples are actually positive.

$$Precision = \frac{\mathrm{TP}}{TP\ +\ FP}$$

Recall

This measure is described as the ratio of true positives to all the positives in ground truth. It explains the model's ability to detect positive samples and in contrast with precision metric, it is dependent only on the positive examples.

$$Recall = \frac{\mathrm{TP}}{TP\ +\ FN}$$

F1-score

It uses a combination of precision and recall. Essentially, it is defined as the harmonic mean of these two metrics. F1-score can often be considered more valuable than accuracy as it takes into consideration the distribution of the data and thus if the dataset is highly imbalanced this metric will provide a better estimation of model's performance.

$$F1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

The values of the aforementioned metrics vary between zero and positive one. The closer the value of the metric to one, the better the model performs. However, in most of the cases, the value of only one metric is not enough to evaluate the total performance of a model and it can lead to misleading results and misunderstanding explanations. Metrics should be used in combination and alongside with other measurements and tools which are the most suit in each case, depending on the nature of the problem.

2.1.2 Linear Regression

Linear regression is one of the most famous and well-studied algorithms in statistics and machine learning, belonging in the category of supervised learning problems. It performs a regression task, in which it models a target prediction value based on independent variables. This kind of models are mainly used for finding out the relationship between variables and doing forecasting. There are a variety of different linear regression models built, which differs on the number of independent variables getting used and the kind of relationship between independent and dependent variables they are considering.

Given an independent variable x (input), a linear regression model learns to linearly map the independent variable x (input) to the dependent variable y (output). In the case of simple linear regression, the target variable is depended only on one input variable and their relationship can be described as a line given by the following linear function:

$$y = a + bx$$

where y is the dependent variable and x is the explanatory variable. The slope of the line is b and a is the intercept (the value of y when x = 0).

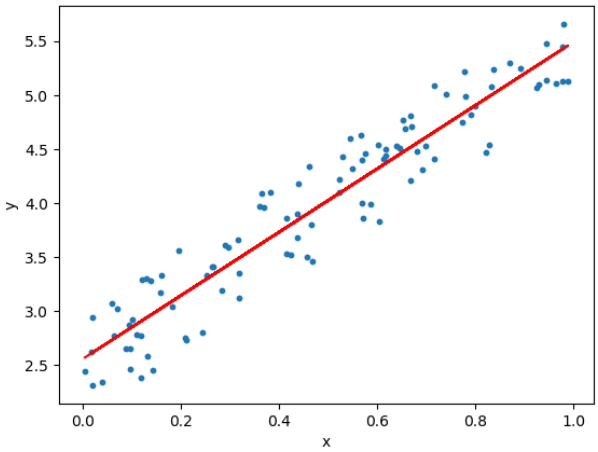

Figure 4: An example of Linear Regression plot

When the target variable is dependent on more than one independent variables, it is defined as multiple linear regression and the relation between inputs and outputs is given by the following equation:

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n = b^T x$$

where $b^T x$ is the inner product between vectors x and b. The b_n parameters are computed using the input variables of the training set and the appropriate ones are those which minimize the error between the predicted value and the real value of the target variable. In Figure 4, an example of a Linear regression plot is given.

2.1.3 Support Vector Machines

Support Vector Machine (SVM) model class was the most popular approach for supervised learning problems, until the arrival of deep learning and neural networks [3]. SVMs build a decision boundary – knows as maximum margin separator, with the maximum possible distance to the example points and, in this way, they are able to generalize well. In this model class, a linear separating hyperplane is created and at the same time the data can be embed into a higher-dimensional space as shown in Figure 5. This can be done using the kernel trick and many times non-linearly separable data in the original input space can easily be separable in the higher-dimensional space.

Additionally, it is worth to be mentioned that SVMs are nonparametric. That means that the linear separating hyperplane is not defined by a set of parameter values but by a collection of example points. An SVM model retain only those examples that are closest to the separating line in contrast with other models such as nearest-neighbor. Thus, SMVs are resistant to overfitting and flexible to represent complex functions as well by combining the advantages of both parametric and nonparametric models. The way that SVMs separate the example points in space is clearly seen in the figure below.

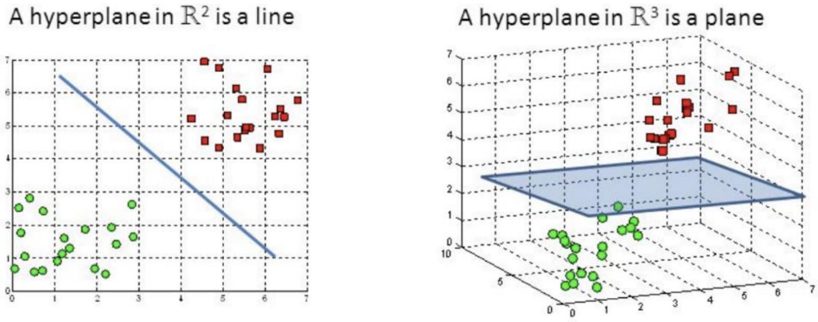

Figure 5: SVM hyperplane

2.1.4 Decision Trees

Decision trees is one of the most popular algorithms in machine learning for supervised learning problems which are used mainly for classification tasks but are able to solve regression tasks as well [4]. They can be easily understood and interpret when visualize which makes them an attractive option for many tasks. A decision tree is a tree-like structure which depict decisions and their possible consequences. In other words, decision trees are used to categorize or make predictions based on how a previous set of questions were answered. Each tree consists of three basic elements: the root node, the decision nodes and the leaf nodes or terminal nodes. From the root node to the leaf nodes flows a series of decisions to be made. Each decision node represents a question or a split point while each leaf node depicts an answer to those questions.

In machine learning problems, the goal of using Decision Trees is to build a training model that will be able to predict the class of the target variable by learning simple decision rules from prior data. In the beginning, the whole training set is considered as the root. The main challenge in decision tree models' implementation is the attributes selection or in other words to select which attributes should be considered as root nodes at each level. Different techniques and measures have been constructed in order to find the most profitable feature for each node with the most famous one being the Information Gain function, which uses the Entropy measure. If a collection of examples has high entropy, that means that there is high heterogeneity between the example of this collection. The entropy is calculated as below:

$$E(S) = \sum_{i=1}^{n} -p_i log_2 p_i$$

Where S is a dataset, n is the different classes of a dataset and p_i the percentage of examples in each class (i=1…n).

The Information gain function computes the entropy's reduction when a split in the examples occurs based on a specific feature. Let S be a dataset and A a feature. The equation of Information Gain is given below:

$$G(S,A) = E(S) - \sum_{v \in V(A)} \frac{|S_v|}{|S|} E(S_v)$$

Random Forest is a famous and extremely important extension of Decision Trees family which can be used both for classification and regression tasks. In both cases, the dataset is used for sampling with replacement to create n different datasets. In the next step, different decision trees are fitted on each dataset and take into consideration only the corresponding features. The output of the random forest is decided by the mean or the average prediction of the decision trees group for regression problems and by the majority of the decision trees ensemble for classification problems.

2.1.5 K-nearest Neighbors

K-nearest neighbors (KNN) is a supervised machine learning algorithm which is used for both regression and classification problems. KNN classifier tries to predict the correct class for a new input based on its k-nearest neighboring examples. In other words, KNN is a distance-based algorithm which main concept is that all the data are placed in a n-dimensional space, where examples that are close are more similar.

When a new example is presented, its distance from each point is computed and the k-nearest points are selected. The selection of k value is an important step and can lead to overfitting to noise. For classification problems, the output class for the new input is the majority class between the k-nearest neighbors and for regression problems the prediction is the mean average of the k-nearest neighbors' values. It is worth to be mentioned that KNN classifiers mostly use the Euclidean Distance metric to calculate the distances between the examples, which is one of the most common measures for distance calculations.

In Figure 6, an example of KNN algorithm is presented. There is a new red input which should be classified in one of the two classes (green or blue). For k=3, the majority class of its neighbors is the green one (2 green vs 1 blue) so the red star should be classified as a green circle. For k=6 the red star should be classified as a blue circle, due to the prevalence of the blue circles in this bigger neighbor (4 blue vs 2 green).

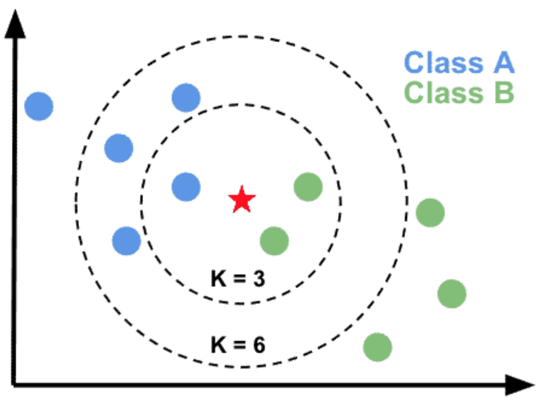

Figure 6: K-nearest Neighbors example

2.1.6 Gradient Boosting

Gradient Boosting is a famous and extremely effective supervised machine learning algorithm and constitutes the basis of the XGBoost library which offers various implementations of this algorithm in many programming languages. XGBoost stands for Extreme Gradient Boosting and it is a distributed, scalable, decision-tree-based ensemble algorithm which provides parallel tree boosting.

Ensemble learning methods combine multiple machine learning algorithms to build a more effective model. Gradient Boosting obtain a better model which consists of multiple decision trees and the key-difference with other algorithms is in how the trees are built and combined. In contrast with Random Forest which uses bagging techniques, Gradient Boosting tries to "boost" or improve a weak model by integrating it with other weak models in order to build a collectively strong model. The idea is different models to be constructed consecutively and each model's performance to affect the construction of the next model. For the classification of new examples, all of the models' decisions are combined through voting with weights, where the famous gradient descent architecture is used. The weight of each model, determines the size of the contribution that this model should have in the final decision and it is proportional to the performance it had on the data it was trained on.

XGBoost is an extremely accurate implementation of Gradient Boosting which is optimized with the application of parallelization in tree building and backward tree pruning. At the same time, this algorithm has been designed to use hardware resources efficient. All the above leads to highly improved model performances and high computational speed which makes the XGBoost the best solution in prediction problems involving small-to-medium structured/tabular data.

2.2 Unsupervised Learning

In Unsupervised Learning, algorithms are trained using information that is not classified or labeled. In other words, there is no a labeled dataset or a known output given to the algorithm and thus the algorithm has to learn without guidance or a supervisor. Here, the task of the machine is to group unsorted information without any prior training by identifying patterns, differences and similarities. In contrast with Supervised Learning, this category of machine learning algorithms is best applied to data that do not have an objective and structured answer. It is more computationally complex and most of the times less accurate than supervised learning methods, even if there is no pre-determination of the correct output for a given input.

Unsupervised Learning has three main categories of tasks: Anomaly Detection, Clustering and Dimensionality Reduction. Clustering is a technique which groups labeled data based on their similarities or differences, often using their distances in space and a distance measure such as the Euclidean Distance. Dimensionality Reduction is a technique which is used in cases where a dataset has too high number of features or dimensions. In such cases there are a lot of risks such as model overfitting and thus the number of data inputs should be converted into a more manageable size. That is what dimensionality reduction methods are doing -most of the times in the preprocessing stage- while preserving the integrity of the data as much as possible. Finally, Anomaly Detection methods, identifies atypical data points, outliers or anomalies within large amounts of data.

The evaluation of Unsupervised Learning methods cannot be as typical and simple as in case of Supervised Learning methods. It depends on the kind of the task and the outputs of each problem. For example, Silhouette Coefficient is a basic metric for Clustering tasks, which used to evaluate clustering algorithms by evaluating the density and the compactness of the new clusters. The score is bounden between -1 and +1 and scores around zero indicate overlapping clusters while scores close to +1 indicate dense, compact and well-separated clusters.

2.2.1 K-means Clustering

K-means is the most commonly used and one of the most famous algorithms for clustering in unsupervised machine learning tasks [5]. It constitutes a centroid-based clustering algorithm which is quite simple to implement and understand and its goal is to group the data into K number of clusters. More specifically, it is an iterative process where each data point is assigned to a group based on its features and in this way data points gets clustered or construct groups with other similar data points. Each cluster can be defined by its centroid and predictions concern the assignment of new examples to the appropriate clusters based on their features. The algorithm can be broken down in the following steps:

1. **Define the number of the clusters** (k) in which the data will be grouped. Choosing the right values of k is one of the most challenging tasks in this problem and it is a critical step for the proper execution of the algorithm. Wrong values of k can affect dramatically the model's performance. Elbow method and Silhouette method are two techniques that can be used to enhance the correct selection of k.
2. **Initializing centroids** by randomly select data points and define them as centroids. Initially, the centroids -which are the centers of the clusters- are selected randomly as they are unknown.
3. **Data points assignment to clusters**. After centroids initialization, data points should be assigned to their closest centroid. For this task, a distance measure is utilized, with the Euclidean Distance metric be one of the most commonly used.
4. **Re-initialization of centroids**. The new centroid of each cluster is computed by finding the average of all data points of each cluster.
5. Steps 3 and 4 are repeated until no more changes occur. That means that all data points have been assigned to the most likely clusters.

As mentioned above, K-means is a very simple algorithm to implement and understand and at the same time it is scalable and fast for huge datasets. It can generalize clusters for different sizes and shapes and new examples' adaptation occurs very frequently. On the other hand, the scalability of this algorithm decreases as the number of dimensions increases and its sensitivity to outliers is quite high.

2.2.2 Hierarchical Clustering

Hierarchical Clustering or Hierarchical Cluster Analysis (HCA) is an unsupervised algorithm which is used for clustering problems. As K-means, this algorithm groups similar objects into distinct clusters which now form a hierarchical structure. It is divided in the following two types:

1. Agglomerative Hierarchical Clustering
2. Divisive Hierarchical Clustering

The Agglomerative Hierarchical Clustering or AGNES (Agglomerative Nesting) is the most common type of hierarchical clustering and it is a bottom-up approach. It starts by considering each point as a cluster and finds the two closest clusters and merge them. This process continues iteratively until a cluster with all the data is created. Divisive Hierarchical Clustering or DIANA (DIvisive ANAlysis Clustering) is a top-down approach which starts with all the data points as a single cluster and then divide the cluster to two least similar groups. It proceeds recursively on each group until there is one cluster for each data point. In both cases, the distance between clusters can be measured with several ways which are known as Linkage methods. Different Linkage methods result to different clusters. The most common Linkage methods are:

- Complete-linkage
- Single-linkage
- Average-linkage
- Centroid-linkage

Hierarchical clustering is a very useful and powerful algorithm for clustering tasks. Compared to K-means, it has the advantage of not having to pre-define the number of clusters that will be created and it performs very well with small datasets. However, it has been characterized as a computationally demanding algorithm which is not effective with vast amounts of data or massive datasets.

2.2.3 Anomaly Detection

Anomaly detection is one of the most common tasks in machine learning field. It is the process of detecting and catching data points in a dataset that deviates from the rest of the data. Also known as outlier detection, anomaly detection is a method used to identify anomalies, atypical data points or unusual behaviors within the data. In particular, as anomaly can be considered every exception and deviation or generally anything that differs from the norm. There are several different types of anomalies such as global anomalies, contextual anomalies, point anomalies, collective anomalies etc. Some common reasons for the emergence of outliers are data preprocessing errors, noise, fraud or attacks.

Irregularities like the aforementioned can be identified by a variety of machine learning algorithms, both supervised and unsupervised. However, the reason that Anomaly Detection concept is mainly intertwined with unsupervised learning is due to the advent of Artificial Neural Networks (ANNs) -which will be analyzed in the next chapter and their extremely effective performance on this task. ANNs can be applied to unstructured data and detect anomalies in unlabeled data and thus decrease the amount of manual work needed in the preprocessing stages. Some of the most common algorithms for Anomaly Detection tasks are the following:

- Local Outlier Factor (LOF) | Nearest-neighbor based
- Cluster based Local Outlier Factor (CBLOF) | Clustering based algorithms
- Decision Trees, SVM, KNN | Classification based

Anomaly detection is a field with rising interest in the scientific community and when it comes to predictive maintenance and monitoring it seems to have a crucial role in the next few years. This came up from the fact that year by year, the amounts of the data produced are vast and businesses should find effective ways and methods to automatically detect anomalies as it is no longer possible to do tasks like this manually. Figure 7 shows an outlier which stands out from the three clusters that have been formed within a dataset.

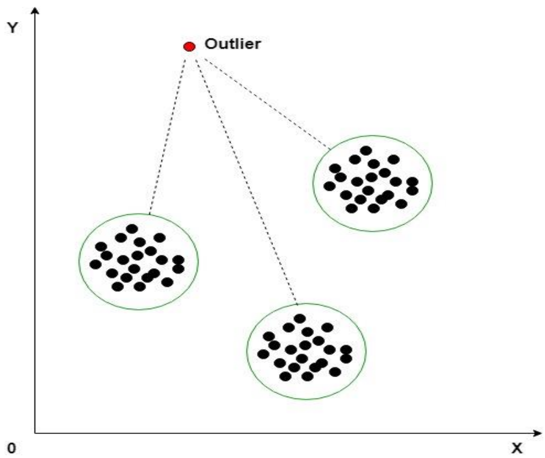

Figure 7: An example of an outlier in a dataset after clustering

2.3 Deep Learning and Artificial Neural Networks

Deep Learning is a subfield of machine learning and it constitutes an extremely novel evolution in Artificial Intelligence science. This type of machine learning imitates the way that human brain obtains knowledge by utilizing the revolutionary technology of Artificial Neural Networks (ANNs), a state-of-the-art approach for building machine learning algorithms and models.

Artificial Neural Networks were inspired by the way that information gets processed and distributed in biological systems or simply in a human brain. Human brain consists of a complex network of billions of biological neurons which works as a computing device by transmitting signals and information and in that way helps human to make decisions very fast. To that end, an Artificial Neural Network consists of several layers with nodes (neurons) which communicate to each other and try to represent knowledge implicitly and cooperate in order to lead to some decisions based on the input data. The term “Deep” in Deep Learning refers to the use of multiple layers in the architecture of an ANN which make it long and narrow or simply deep [6].

As shown in Figure 8 an ANN can consist of several layers of neurons. The first layer is always called input layer. Input layer -as its name suggests is the first layer in which the input data are imported either in batches or all at once. In the same way, the last layer in an ANN is the output layer and it gives the final results or simply the output data of the network. Between input and output layers, there may exist hidden layers depending on the task. In this architecture, there are edges which connect the neurons of each layer or simply the layers to each other. The inputs can be changed from layer to layer, using weights which are positive or negative and can be adjusted through the training process of the ANN. It is worth mentioning that each neuron consists of the sum or the aggregation of the weighted inputs received of this neuron and some activation function which will be described in the next subsection. The final output of a neuron can be used as input for another neuron.

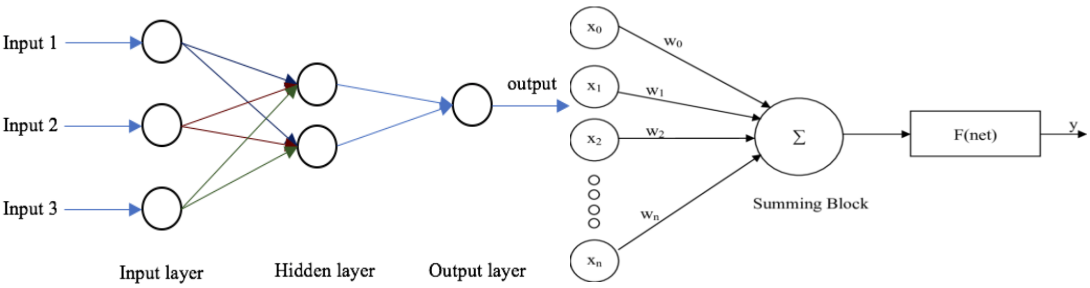

Figure 8: A simple ANN architecture and the anatomy of a neuron

In contrast with the common machine learning approaches, deep learning algorithms can ingest and process unstructured data and thus eliminate some of the preprocessing work. Additionally, deep learning models are capable of several types of learning such as supervised learning, unsupervised learning and reinforcement learning. The latter is a process where the model becomes better and more accurate by taking feedbacks for the action that performed and trying to maximize the reward.

2.3.1 Activation Functions

Activation or threshold function *f* is a part of every neuron in an Artificial Neural Network and can be described as a filter that forms the final output of the neuron. It is a critical part in the phase of designing every neural network as they define how the weighted sum of the inputs will be transformed into the output of a node (or nodes) in a layer. The selection of the threshold function is crucial in order to train a sufficient model as it has a large impact on the performance and the capability of the neural network.

Different layers can use different activation functions. More specifically, all hidden layers use the same activation function and the output layer will typically use a different function which is dependent on the type of prediction that the task require. There is a variety of activation functions used in neural networks but only a small number of them are used for the most of the tasks in hidden and output layers. As can be seen in Figure 9, activation functions can be divided into two basic types: Linear and Non-Linear functions.

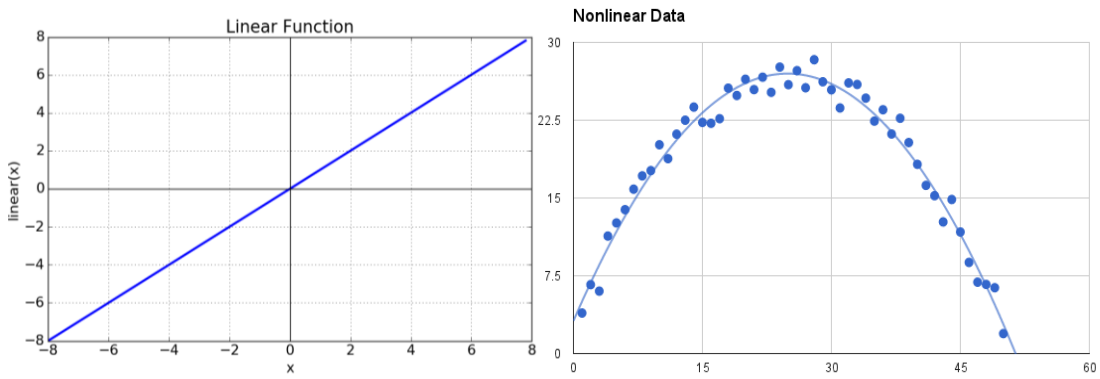

Figure 9: An example of a linear (left) and a non-linear (right) function

The basic functions that should be considered for use in hidden layers are the following:

- Rectified Linear Activation or **ReLU** is probably the most common function used for hidden layers and it is quite simple and effective. It can be calculated as follows:

  $$\text{relu(x)} = \max(0, x)$$

  This means that the value 0 will be returned if the input value x is negative. Otherwise, the value x is returned.

- Logistic or **Sigmoid** activation function is the same function used in the Logistic Regression Classification algorithm. It takes as input any real value x and outputs values between 0 and 1. It is calculated as below:

  $$\text{sigmoid(x)} = \frac{1.0}{1.0 + e^{-x}}$$

  where e is a small mathematical constant.
- Hyperbolic Tangent or **Tanh** is very similar to the sigmoid function, and it has a similar S-shape. It takes as input any real value x and outputs values between -1 and 1. It is calculated as follows:

  $$\text{tanh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

  where e is a small mathematical constant.

The most commonly used functions for output layers are the following:

- **Linear** activation function is also known as “identity” or “no activation”. This is because it just returns the weighted sum of the inputs directly without to make any changes on them by multiplying it by 1.
- Logistic (Sigmoid)
- **Softmax** function is a “softer” version of argmax that gives a probability-like output. In other words, it outputs a vector of values that sum to 1 and thus they can be considered as a set of probabilities. Softmax is calculated as below:

  $$\text{softmax(x)} = \frac{e^{x}}{\text{sum}(e^{x})}$$

  where e is a small mathematical constant.

2.3.2 Feed-Forward Neural Networks

The Feed-Forward Neural Network (FFNN) was the first type of neural network and additionally the simplest to understand and implement. In this type of neural networks, information flows in only one direction, from the input layer through the hidden layer -if any and to the output layer. Connections between neurons cannot form a cycle or loop. The simplest neural network which falls in this category is known as Single-layer Perceptron and is often used in classification tasks. This network consists only of one layer. A set of inputs enter the layer and get multiplied by the corresponding weights. Then, the sum of the weighted inputs is calculated. If the sum falls below than a specific threshold (most of the times zero), the output of the network is -1, whereas if the sum is greater than the threshold, then the output is 1. This is depicted in Figure 10.

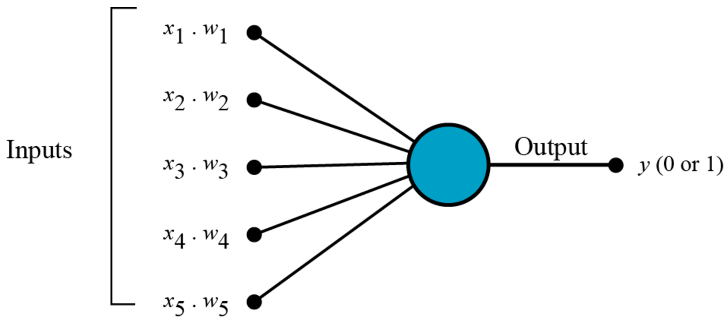

Figure 10: A single-layer Perceptron

Multi-layer perceptron (MLP) is another class of artificial neural networks which are usually interconnected in a feed-forward no-cycle way. A multi-layer perceptron model consists of multiple layers of neurons and each neuron in one layer has one-way connections to the nodes of the next layer. MLPs should have at least three layers of nodes among which an input layer, a hidden layer and an output layer. Nonlinear activation functions are used among the nodes of such a network. Figure 8 which was presented at the beginning of the section is a simple example of a multi-layer perceptron architecture, which consists of exactly three layers.

2.3.3 Gradient Descent

Gradient descent is an algorithm which is commonly used to optimize the performance of machine learning models and mainly neural networks [7]. During models' training, the cost function of gradient descent acts as a barometer by judging their accuracy and updating their parameters iteratively until to achieve the best performances. A cost function close or equal to zero means the smallest possible error and only then the parameters stop being adjusted. The gradient descent algorithm behaves very similar to linear regression but it is based on a convex function as it can be seen in the figure below.

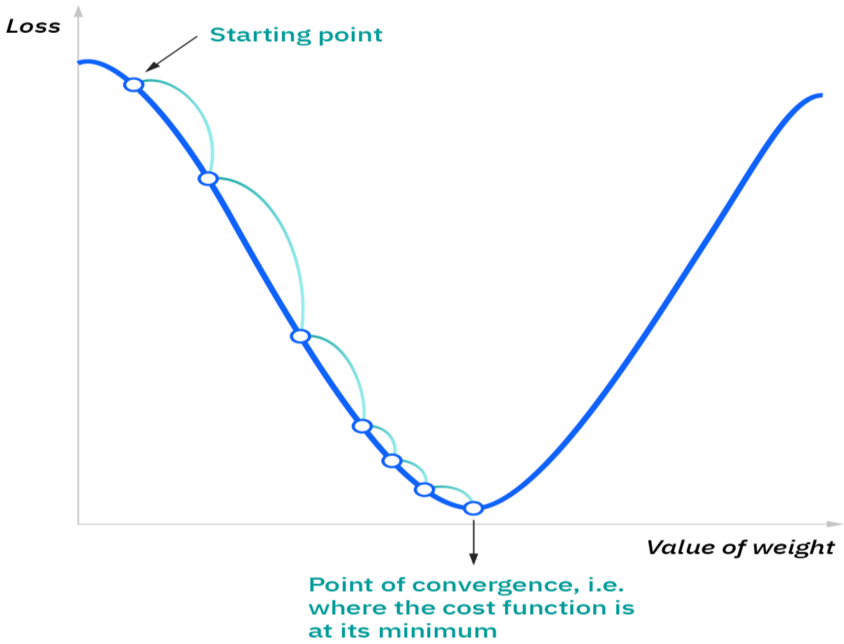

Figure 11: The loss function of gradient descent

It starts from an arbitrary point just to evaluate the model's performance and from there it will find the slope and then use a tangent line to observe the declivity of the slope. The updates of the parameters are informed with the help of the slope. Specifically, the slope at the starting point is steeper and as the parameters get adjusted, the steepness starts gradually to reduce until it reaches the point of convergence or the lowest point on the curve. The goal of the gradient descent algorithm is to reach this point of convergence which equals to minimizing the cost function or the error between the predicted and actual values.

A direction and a learning rate are needed for the aforementioned task. Learning rate or step size is the size of the steps that should be taken in order to reach the point of convergence and it is based on the behavior of the cost function. Loss function or cost function gauges the error between the predicted values and the actual values and gives feedback to the model in order to adjust its parameters and minimize the error by finding the local or global minimum (point of convergence) which is shown in Figure 11. Successive iterations take place with a direction to the steepest descent until to loss function get close or equal to zero. There are three different types of gradient descent algorithms:

1. Batch Gradient Descent
2. Stochastic Gradient Descent
3. Mini-batch Gradient Descent

2.3.4 Back Propagation

Back propagation is an algorithm used in machine learning tasks which is based on calculations of the gradient of the loss function and mainly used in neural networks. Gradient descent and back propagation are two different methods that form a strong combination which enhances the learning process of neural networks. The back propagation algorithm is probably the most fundamental building block in every neural network and is used to train them via the chain rule method. It is named so because the weights are updated backward, from the output to input. During training, every forward pass in a network is followed from a backward pass in which the model's parameters (weights and biases) are tuned. In other words, this method takes the neural network's output error and propagates it backwards to the previous layers. In this way, back propagation identifies which pathways within the network are more influential in the final prediction. Then, it strengthens or weaken the corresponding connections in order to lead the network towards the desired predictions.

The advantages of using this algorithm are several as it is a standard process that usually works well. There aren't any parameters to be tuned except the number of the inputs and it constitutes an efficient and adaptable algorithm that does not require any prior knowledge for the network. On the other hand, back propagation performance is highly dependent on the input data and its training process is time and resource-intensive. Additionally, this algorithm is sensitive to noise and irregularities and it uses a matrix-based approach instead of mini-batch approach which is a computationally less efficient process. There is a variety of different algorithms that implement back propagation which are used to optimize the training process of a neural network and for this reason they are known as optimizers. Adam optimizer seems to be one of the most commonly used algorithms as it is able to achieve fast convergence and to train complex and deep neural networks efficiently.

2.3.5 Normalization

Normalization is a preparation technique through which the data are transformed in order to be on a same scale. It is useful when the distribution of the data isn't known and the data has variable scales as it can improve the training stability and the performance of the model. In the case of artificial neural networks, where assumptions about the distribution of the data are not made, normalization techniques are extremely important as they can help the training process of the network. This happens by transforming the different features of the networks so as to be on a similar scale and in this way to stabilize the gradient descent step and allow larger learning rates to be used, allowing the network to converge faster. Several normalization techniques have been implemented and the selection of the most proper depends on the task each time. Four of the most commonly used are the following:

1. **Scaling to a range**
   In this technique, scaling means to convert feature values from their natural range into a standard range (usually -1 to +1 or 0 to 1) using the formula below:
   $$X = \frac{x - x_{min}}{x_{max} - x_{min}}$$
   It is recommended when there are no or few outliers on the data and the approximate upper and lower bounds are known.

2. **Feature Clipping**
   When the data contains extreme outliers, feature clipping seems to be an appropriate solution as it transforms all the features above or below a certain value to a standard value. For example, it could clip all the temperature values below -20 to 0.

3. **Log Scaling**
   It computes the log of the values in order to create a narrow range of them and it is useful when the data follows the power law distribution as it changes the distribution and helps linear models improve their performances.

$$x' = log(x)$$

4. **Z-Score**
   This technique is useful when there are a few outliers but not so many that feature clipping is needed. It represents the number of standard deviations above or below the mean a data point is and can be both negative or positive. It is calculated with the following equation:

$$x' = \frac{x - \mu}{\sigma}$$

2.3.6 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is a type of artificial neural network which became dominant in several image classification and object recognition tasks, which require the ability to identify patterns and make sense of them [8]. CNNs have three main types of layers which distinguish them from other types of ANNs: Convolutional Layer (input layer), Pooling Layer (hidden layer) and Fully-connected layer (output layer). Additionally, they are designed in a way that allows them to automatically learn spatial hierarchies of features by using the aforementioned blocks and backpropagation techniques. The nodes' connections are quite similar with a simple feed-forward ANN with the difference being that in a CNN architecture, a node is activated and send its output to the next layer only if the output value is above the predetermined threshold value.

The core building block of a Convolutional Neural Network is the convolutional layer as the majority of computation occurs there. It consists of three parts which are the input data, a kernel or a filter, also known as a feature detector and a feature map. The process of checking if a feature is present in an image by moving across its receptive fields is called convolution and it is executed by the feature detector, a two-dimensional array of weights which represents a part of the image. A series of dot products between the filter and the input image pixels are calculated and the final result is known as a feature map or activation map. By this process, CNNs achieves a compressed representation of the input by reducing the dimensionality and keeping only the necessary information for the task.

The Pooling layer which follows the convolutional layer is utilized for further dimensionality reduction. Similar to the convolutional layer, it sweeps a feature map across the input but in this case, the filter does not have any weights. Max pooling and average pooling are the two main types of pooling. Finally, the fully-connected layer or the output layer of a CNN consists of nodes that each of them is directly connected to a node in the previous layer. In the fully-connected layer, the classification task takes place based on the extracted features which emerged from the filtering in the previous layers. An example of a Convolutional Neural Network architecture is shown in Figure 12.

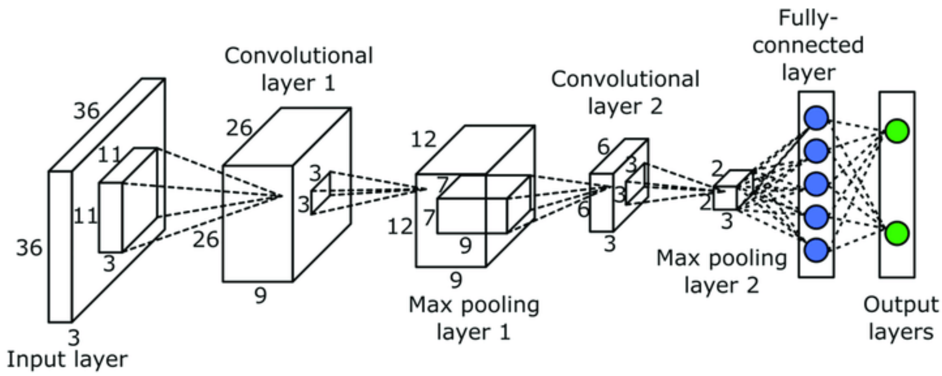

Figure 12: A CNN with two convolutional layers

2.3.7 Recurrent Neural Networks

A Recurrent Neural Network (RNN) is a type of artificial neural network where the neurons in one layer have connections with neurons of previous layers and is commonly used for tasks with sequential data such as text or time series data. In other words, the main difference between RNNs and feed-forward neural networks is that the first have connections that point backwards. Elman was the first to introduce the base recurrent unit in 1990 [9]. The main attribute of RNNs is that every neuron has a “memory” mechanism in order to remember the output of previous layers which influence the current input and output each time. That means that the output of an RNN depend on the prior elements within the sequence. In contrast with FFNNs, RNNs share same weight parameters across each layer of the network which are adjusted through the backpropagation and gradient descent processes. A simple RNN architecture is given in Figure 13. There are several variants network architectures of RNNs with two of the most dominant ones being the following:

- **Long Short-Term Memory (LSTM)**
  The LSTM constitutes an improvement of the Recurrent Neural Network which address the failure to address the problem of vanishing gradient descent. This problem occurs in the training phase of a neural network, when the gradient starts getting exponentially small in size so that the updates of the parameters become insignificant. As a result, the neural network is unable to learn. LSTMs remedy this problem by having cells in the hidden layers which consists of three gates: an input gate, an output gate and a forget gate. These gates keep only the information needed to predict the output in the network and forget the rest and thus helping the gradient size to remain constant. For this reason, LSTMs are considered an ideal solution for problems which consists of long-term patterns in data and long dependencies such as time series and texts [10].
- **Gated Recurrent Units (GRUs)**
  Similar to LSTM, GRUs is a variant of RNNs which addresses the short-term memory problem [11]. Instead of the cell state of LSTMs, GRUs use hidden states with two

gates: a reset gate and an update gate which regulate the information flow. This makes GRUs less complex than LSTMs, easier to modify and faster to be trained. LSTMs outperform GRUs in tasks where modeling long-distance relations is required, while GRUs are computationally more efficient due to their structure and more effective in small datasets.

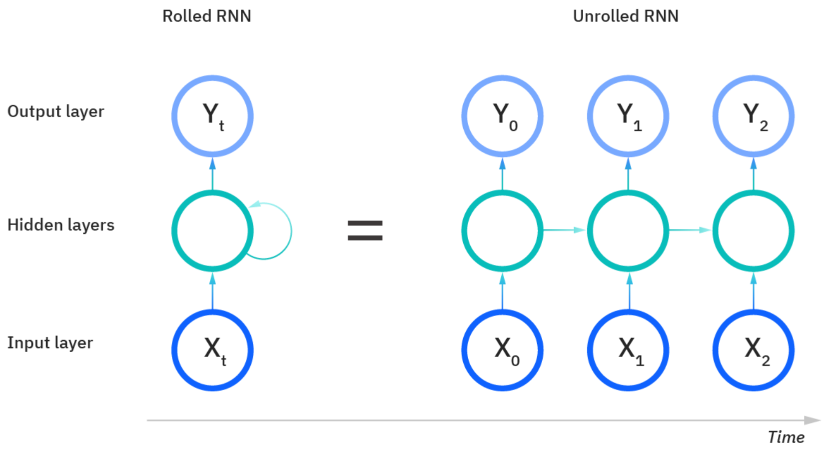

Figure 13: A simple RNN in both rolled and unrolled views.

2.4 Time Series Forecasting

Time series forecasting is the process of analyzing time series data in order to make predictions and facilitate decision-making. Any kind of data, that are on the same subject and can be captured at different time intervals can be considered as a form of time series data. Based on the assumption that future trends will be similar to historical trends, future events can be predicted by analyzing the trends on past data. Forecasting on time series data is not always exact predictions but it can be likelihood forecasts depending on the data. It is worth to mention that statistics and modeling tools are used across all the scope of time series forecasting in order to analyze the data, extract their statistical properties and make forecasts, with machine learning entering the field the recent years with significant impact.

A large volume of data is always required in this task, in order to ensure the best possible results which would be reliable and consistent, otherwise, noisy data and outliers will lead the models to misleading forecasts. Timeseries forecasting has a range of applications in several industries such as weather forecasting, finance, energy demand, healthcare etc. [12] and constitutes a vital tool for most of the companies which follow long-term growth strategies. In Figure 14 an example of time series dataset can be seen. The dataset concerns the furniture sales of a store within seven years.

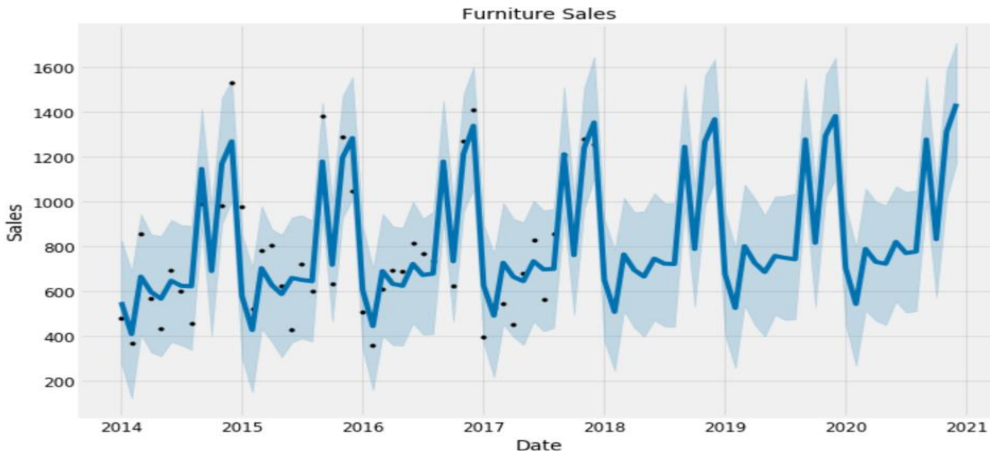

Figure 14: Furniture sales time series dataset

2.4.1 Decomposition of Time Series

Decomposition of time series involves thinking of a series as a combination of some properties or components and provides an abstract model of analyzing time series generally during forecasting which enhance the prediction process. The main properties of a time series are presented below and can be seen visualized in Figure 15.

- The **trend** (T_t) component at time *t,* which shows an ascending or descending pattern in the data throughout the time. The trend component does not have to be linear.
- The **cyclical** (C_t) component at time *t,* which reflects repeated fluctuations in the data which are non-periodic. These fluctuations are not seasonal and their duration depends on the nature of the time series. Cyclical component is often grouped up with trend.
- The **seasonality** (S_t) at time *t,* which describes seasonal patterns or repeating cycles within the series which occur at fixed time-intervals. Seasonality exists only when a time series is influenced by seasonal factors.
- The **residual** (R_t) component at time *t,* indicates random and unpredictable influences in the data which are not repeated in regular intervals. Residual can be considered anything that does not fit to the other three main components.

Time series X in time *t* can be modeled with two different ways depending on the problem:

1. The additive model: $X_t = T_t + S_t + R_t$
2. The multiplicative model: $X_t = T_t \cdot S_t \cdot R_t$

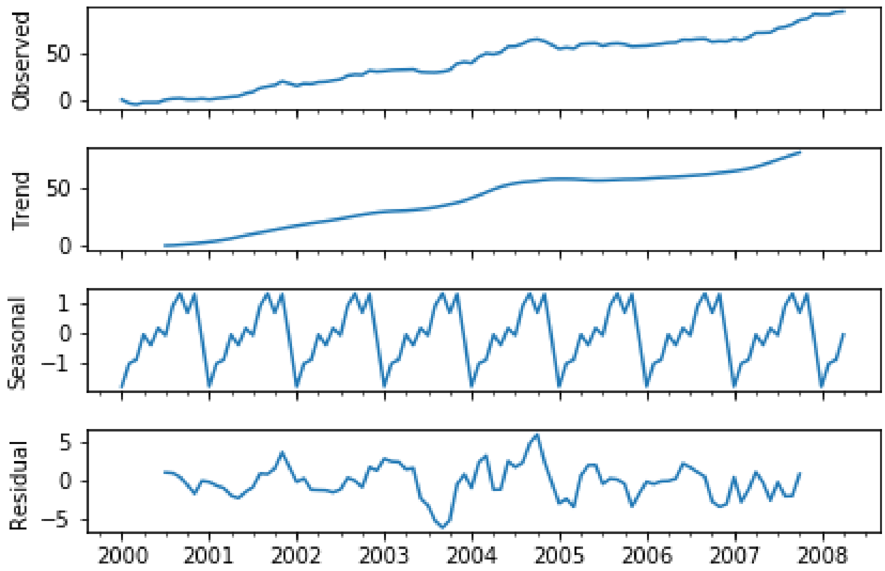

Figure 15: An example of time series decomposition

2.4.2 Statistical Models for Time Series Forecasting

The decomposition of a time series follows a few predetermined - usually - steps. First, the trend is detected and removed. Next, if present, one or more seasonalities are extracted. The remainder is the residual component. For the execution of the aforementioned steps several statistical models and techniques have been developed such as the **differencing method** [13], linear and nonlinear regression methods, the **X-11** method [14] and the **Moving Average Smoothing** (MAS) [15]. The most widely used statistical approaches for time series forecasting are **Exponential Smoothing** and AutoRegressive Integrated Moving Average or simply **ARIMA** models. As later analyzed, ARIMA models aim to describe the autocorrelations in the data, while Exponential smoothing models are based on the description of the seasonality and trend in the data.

More specific, Exponential Smoothing (Brown 1959; Holt 1957; Winters 1960), is a statistical technique which was proposed in the late of 50s and has motivated some of the most successful forecasting methods that are based on it. The assumption behind the forecasting methods that use Exponential Smoothing is that the more recent an observation is, the more related with the current observation. In other words, observations which are temporally close, have similar values as well. The forecasts produced from Exponential Smoothing methods are weighted averages of past observations, with the weights getting reduced in an exponentially manner as the observations get older. Several methods based on Exponential Smoothing have been developed: some of are based on trend such as Holt's linear trend method, while others are based on seasonality such as Holt-Winters seasonal method.

The selection of the appropriate forecasting method depends on the characteristics of the time series data each time and the way in which the data can get into this method (additive or multiplicative model). Where trend and seasonality patterns are not clear in a dataset, Simple Exponential Smoothing (SES) is used. SES follows the aforementioned concept where weights which are associated with older observations are smaller. A one-step-ahead forecast is given by the following equation:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots$$

where α is a smoothing parameter between 0 and 1. According to the previous equation, the one-step-ahead forecast for time T+1 is a weighted average of all observations in the time series $y_1, \dots, y_T$ with the weights be decreased by a rate which is controlled by the parameter α.

On the other hand, ARIMA models are related and used in stationary series. **Stationary** time series are those whose properties do not depend on the time at which the series are observed and they look similar at any time point. Time series with trend and seasonality components cannot be considered stationary. **Differencing** is a method by which a non-stationary time series is transformed to stationary by computing the differences between consecutive observations. ARIMA models consist of the three following components:

- **Autoregressive** models which assume that the output forecast variable is a linear combination of the past values of the variable. An Autoregressive model AR(p) where p is the past values taking into consideration is given by the following equation:
$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$
Where c is a constant, ε_t is an error term know as white noise which count the errors and $\phi_1, \dots, \phi_p$ are the parameters of the model.
- **Moving average** models which consider the forecast value as a linear combination of the previous q error terms. A Moving Average model of order q or MA(q) can be calculated with the formula below:
$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$
Where c is a constant, ε_t is white noise and $\theta_1, \dots, \theta_q$ are the parameters of the model.
- **Integration** of the above which are the combination of both past values and past error terms. This combination is modeled on the differenced time series and gives the final ARIMA model. Based on the previous equations for a d-order differenced series, an ARIMA (p, d, q) model is given by the following formula:
$$y'_t = c + \varphi_1 y'_{t-1} + \dots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

2.4.3 Machine Learning for Time Series Forecasting

Despite the existence of several remarkable and highly efficient statistical models for forecasting, the goal of this thesis is to use and examine machine learning techniques and approaches in time series data in order to do forecasting and evaluate the results. As every machine learning task, the time series data should be transformed in a clear dataset with specific features and attributes. The difference between time series problems and classic machine learning tasks is that the first use data which are sampled based on a timed-based dimension. Regarding the DateTime attribute, a variety of different time features can be created by dividing it in subcomponents such as days, weeks, months, quarters of year or year of an event.

Additionally, the training and test sets should be created by splitting the data. This can happen with the classic method of splitting the dataset into two parts or with the sliding window technique. More specific, in this technique a small window of time (time step) is used as input data to train the model and then the model is tested in the next window which is used as the output. The forecast horizon is the set of points in the future that will be predicted from the model and should be defined. In Figure 16 an example of sliding window method is shown. The forecast horizon has 20-days (N) length and in every step a forecast output for the next day is given by the model. The window rolls to the data and the model is able to make predictions of up to N-1 days as can be shown in the figure below.

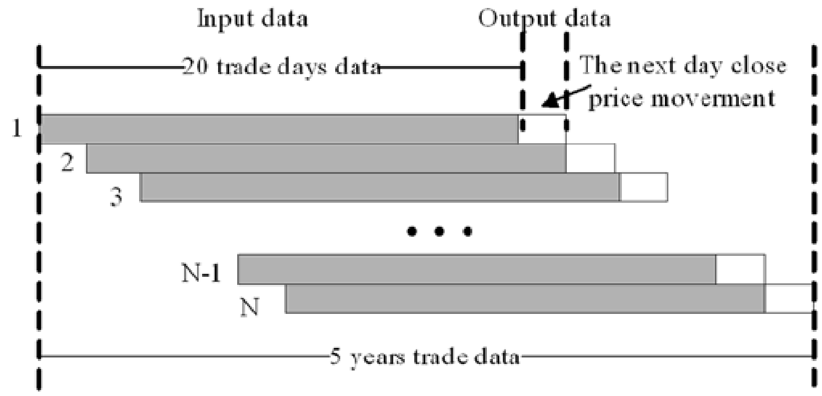

Figure 16: Time series sliding window method

In both cases of splitting a time series dataset into training and test set, it is essential that the order of the data does not change after the split. In more formal terms, a machine learning forecasting task can me modeled as a function f which receives multiple sequences x_i and outputs a single sequence y. The equation is given below:

$$\boldsymbol{y} \in \mathbb{R}^{H,p_h} = [y_{T+p_h+1}, y_{T+p_h+2}, \dots, y_{T+p_h+H}]$$

$$X = \boldsymbol{x}_i \in \mathbb{R}^{W,p_w} = [y^i_{T-p_w-W}, \dots, y^i_{T-p_w}], for\ i = 0, \dots, N$$

$$\widehat{\boldsymbol{y}}_{T+p_h \mid H,p_w} = \mathcal{F}(X)$$

or:

$$\begin{bmatrix} y^0_{T-p_w-w} & y^0_{T-p_w-w+1} & \cdots & y^0_{T-p_w-1} & y^0_{T-p_w} \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & \\ y^N_{T-p_w-w} & y^N_{T-p_w-w+1} & \cdots & y^N_{T-p_w-1} & y^N_{T-p_w} \end{bmatrix} \quad \mathcal{F} \quad [y_{T+p_h+1}, y_{T+p_h+2}, \dots, y_{T+p_h+H}]$$

where W is a predetermined history window, H is the horizon which takes possible offsets, p_w the history offset, p_h the future offset and x_0 the historical values of y.

It should be noted that visualization is an integral part of time-series analysis and forecasting. Visualizing time series data can help to detect patterns and their components (trend, cyclical, seasonality, residual), the existence of correlations between the variables and if the data is stationary or not. It seems that several machine learning algorithms such as Decision Trees, SVMs, KNN, XGBoost etc. work well with time series data. However, Artificial Neural Networks like LSTMs are proven to have the best results working with such type of data since they have the ability to remember and use previous information to predict every next output. In Figure 17 a visualization of a probabilistic forecast given by a machine learning algorithm is given.

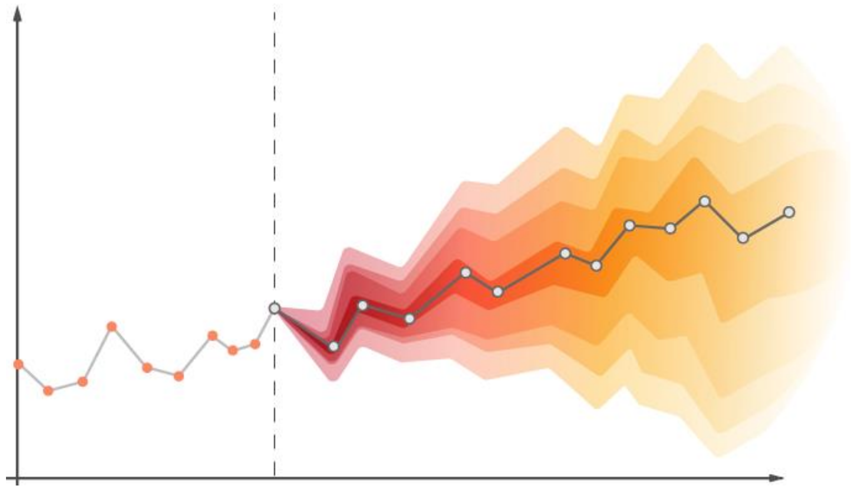

Figure 17: A visualization of a probabilistic forecasting using ML model

Chapter 3

Renewable Energy

The energy itself constitutes a vital aspect in human's life for the last centuries. The evolution of humanity is closely related with the use of energy. Fire was discovered by the human about 500.000 years ago and was used for protection, warmth, cooking and lighting as one of the first sources of energy harnesses by human. Water and wind were introduced some thousands of years later and exploited for several tasks such as agriculture and boating. Todays' societies and the global financial system is based in a massive energy market and the utilization of the energy through various ways for the satisfaction of every kind of human needs. The most common energy product is electricity which is used for the most of the daily tasks in human societies such as heating or even transportation. At the same time, conventional energy sources which produce the aforementioned energy products has not only started to eliminate, but they are extremely harmful for the environment as well, as it will be analyzed in Section 3.2 of this chapter. Figure 18 shows the global energy consumption from 1965 to 2019 [16].

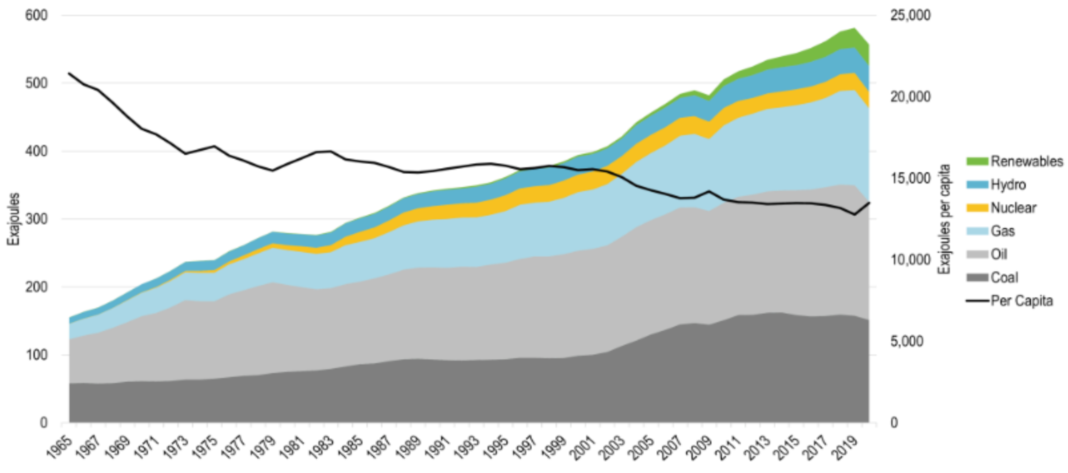

Figure 18: World energy consumption, 1965-2019

According to recent research (2020), the main energy sources used within the European Union members in descending order are: oil and petroleum products, natural gas, nuclear energy and solid fuels [17]. The energy available of each country consists of different combinations between the aforementioned energy products. For example, in Italy, about 80% of the available energy comes from gas and oil, while in Germany the same percentage is distributed between oil, gas and coal [18]. Additionally, different sectors of economy, requires different types of energy. Thus, each country's energy mix is based on factors like industry size, climate, transportation use by citizens etc. Energy "lifecycle" starts with its production, continues with transformation to energy products (electricity, cooling, heating, natural gas, diesel etc.) and with the distribution of them and ends with their consumption. It is worth to mentioning that the energy available in EU derives both from energy produced by EU members (42%) and mainly from energy imported from third countries (58%) [19].

As conventional energy sources seem not to be a viable solution, one of the biggest challenges of this century is to find and use efficient and sustainable energy sources to fulfill these needs. Thus, the renewable energy came to the fore and organizations, governments and the scientific community all around the world have focused on finding and utilizing alternative energy sources which will ideally replace the conventional ones. **Renewable energy** or **green energy** is energy produced from natural sources such as sun, rain and wind which are plentiful and replenished faster than they are consumed.

In contrast with conventional energy sources such as coal, oil and fossil fuels, the utilization of green energy sources does not require any active intervention such as mining, pumping or burning but just the utilization of the already existing flow of energy in nature. The term "renewable" is not completely representative for any type of green energy source as some of them need some thousands of years to renew themselves. A great advantage of renewable energy sources is that they are widely available, almost in every geographical area due to their dependence on natural phenomena. Furthermore, as said above it is more sustainable for the environment and endless, more flexible concerning their production and extremely cheaper than the conventional energy sources. As shown in Figure 19, Renewable energy sources constitutes 17.4% of the total energy available in EU [20].

However, green energy faces some disadvantages that makes its predominance in relation with conventional energy sources doubtful. Firstly, they are not as efficient as the conventional ones. Actually, their coefficient of performance is significantly low (30% or even less) and thus the overall cost of constructing a renewable energy production unit is quite high as large areas of land are needed. Additionally, their performance is highly dependent on weather conditions which could make the amounts of energy produced unstable and unpredictable. Finally, in order to overcome the previous issue highly effective energy storages are necessary which makes the total process even more difficult from a financial aspect mostly when it comes for large-scale energy production units.

Due to the aforementioned reasons, it is considered extremely difficult, even impossible for the renewable energy sources to be used exclusively and replace the conventional sources in order to satisfy the needs of the large urban centers. Although, as it will be shown in the next section, the non-negotiable need for a sustainable future makes the research society to focus on finding ways to make renewable energy more efficient. The current thesis examines and presents one of them, which is the forecasting of renewable energy units' productivity.

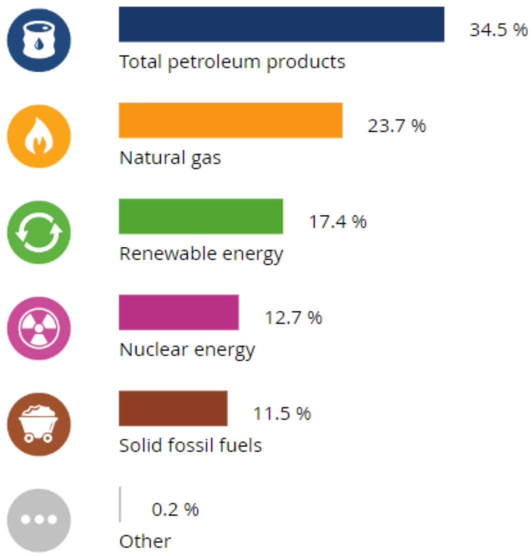

Figure 19: Energy mix for the European Union

3.1 Renewable Energy Sources

Renewable energy can be produced widely in many geographical areas in contrast with conventional energy production which is concentrated in a certain number of places and countries. They are based on the environment's natural phenomena and they are both sustainable and unlimited. There are several types of renewable energy sources with the officially recognized and commonly used ones being the following: solar energy, wind energy, hydropower energy, biomass energy, marine energy and geothermal energy. There are also energy sources such as nuclear energy for whom it is controversial if they can be considered as green sources. Some of the previous energy forms of energy such as marine and geothermal energy are produced from natural phenomena such, while others produced directly (solar energy) or indirectly (wind, biomass) from the sun. In this chapter, the six main types of renewable energy sources will be analyzed and the controversial nuclear energy will be discussed.

3.1.1 Solar Energy

Solar energy is energy produced by the sun and converted into electrical or thermal energy. It constitutes the cleanest and the most promising renewable energy source because of its abundance and its nonpolluting character. It has various forms as it can produce heat, generate electricity, provide light or cause chemical reactions. It is said that with the construction of a specific number of solar energy collection units and their suitable exploitation it is possible to cover the entire planet's energy needs.

Solar energy has two forms that can be used, first the sun's heat and secondly the sun's light. Solar energy collection units are divided into two categories based on the form of solar energy that they use. The first group is solar thermal energy systems which utilize the sun's heat and produce electricity, while the second category known as photovoltaic systems, exploits the sun's light through the photovoltaic effect in order to convert solar radiation into electricity as well. In simple terms, the photovoltaic effect is a process that generates voltage or electric current in a photovoltaic cell when it is exposed to sunlight.

Photovoltaic systems are one of the most exploited and widely used applications for renewable energy production. They consist of one or more panels and each panel consists of photovoltaic cells, an inverter and several electrical components which makes them capable to produce electricity through sun's light. Each cell has a length between 12 and 16 cm, a square shape and it is made of semiconductor materials. These materials, when exposed to sunlight radiation, creates polarization of electrical changes. Thus, an electrical potential difference between two poles is created or simply an electrical generator on which the photovoltaic effect is based on.

An example of photovoltaic panels is shown in Figure 20. A photovoltaic system does not only consist of photovoltaic panels. These panels are also connected with several devices and tools such as trackers, batteries, inverters, monitoring etc. which helps to convert the electrical energy produced in the appropriate and desired form. The power produced by a photovoltaic system can vary between a few watts to hundreds of megawatts, depending on the size and the technologies behind the system.

Figure 20: Photovoltaic panels

As mentioned before, photovoltaic panels are extremely commonly used and have several applications. They can be placed on buildings' rooftops, where they are able to maximize the light absorption in urban and inhabited areas or feed the buildings with electricity which is produced in distant from the building locations via their connections. They are able to support and supply local users and homes or even to completely cover the energy needs of commercial areas. Figure 21 shows an example of building embedded photovoltaic system. Photovoltaic systems of a bigger power capacity or solar farms are large groups of photovoltaic panels placed in vast areas and often used and exploited by huge renewable energy corporates. Additionally, there are other types of photovoltaic systems installations such as those mounted in vehicles which are used to supply applications with smaller energy requirements such as energy phones or lamps. Photovoltaic systems have even been used to power solar spacecrafts.

Figure 21: Building embedded photovoltaic system

The main factors that can affect the efficiency of a photovoltaic system are the climate and meteorological elements of an area as expected and also the position of the system. Factors such as the orientation and the tilt of the panel's installation, the humidity and the intensity of the sunlight and heat are strongly linked with the efficiency of a photovoltaic panel. Extremely high humidity or temperature can cause significant drops in the efficiency of a photovoltaic panel.

Concerning the way that a photovoltaic panel should be placed and installed, there are specific directions in order to be able to maximize its efficiency by utilizing the incident solar radiation in an optimal way. The first step is always to review the area where the panel will be placed. Large buildings, trees or other structures around the installation could limit the sight of line with the sun and thus its efficiency. Sometimes panels are placed horizontally to face the sky directly, while others are placed with specific direction and selected angle. In the latter case, the photovoltaic panel should always face toward the equator. For instance, a photovoltaic panel in San Diego, California, should be placed with an angle of approximately 33° and face south. These directions are general recommendations which takes into account the angle of the sun over the year. To maximize the efficiency of a photovoltaic panel in winter and mainly in case of northern latitude areas there is a general rule. Panels should be tilted up from horizontal at an angle approximately 15° greater than the latitude.

The use of photovoltaic systems has many advantages which makes them one of the most promising renewable energy solutions for the future. Solar energy is unlimited, everywhere accessible and costless. Photovoltaic systems and installations are quiet when they work, resistant to time and adaptable when it comes to how they will be placed. However, their dependence on unpredictable factors such as the availability of sunlight constitutes a major disadvantage. Also, the cost of their installation remains extremely high which makes them unaffordable, while supporting storage and backup tools are always prerequisites to such installations.

When it comes to the exploitation of the sun's heat for energy generation, solar thermal systems are the most widely used solution and constitute the second basic group of solar energy-harvesting systems. As mentioned before, solar thermal energy systems convert solar radiation into thermal energy which can then generate electricity. There are three types of solar thermal systems, also known as collectors depending on the volume of the heat that they are able to produce: low, medium and high temperature collectors. Low and medium temperature collectors are not used for electricity generation but mainly for water heating in industrial and domestic infrastructures. This happens through the flat plates which constitutes the main installation of a collector and they are able to capture solar energy and use it to heat water inside a frame. A common and widely used example of this category of solar energy-harvesting systems is the solar water heater which is shown in Figure 22.

Figure 22: Solar water heater

On the other hand, high temperature collectors are more complex and used for more energy demanding tasks, mostly in solar thermal plants. They are consisting of concentrating solar power technologies which use mirrors or lenses to concentrate sunlight through a receiver. The most common type of receiver is the water tank and it is where the sunlight is collected with the aim of mirrors or lenses. In simple terms, the heat from the sun is transferred to the water, which then get heated and converted to gaseous form (steam). Finally, a steam generator is used which converts the steam into electricity. High temperatures of the source can improve the efficiency of heat engines. Depending on the type of receiver, mirrors or lenses they use, there have been developed several concentrated solar energy technologies. The most well-known and commonly used are parabolic trough systems, power tower systems and dish/engine systems. In Figure 23 an example of a parabolic trough system is given.

Despite the high energy storage needs, thermal energy constitutes a more cost-efficient solution than conventional energy sources, as heat storage is much cheaper and more efficient than electricity storage. It can be considered as a space efficient technology and totally green energy source as it doesn't harm the environment and it is infinite. However, like the photovoltaic energy systems, it is highly dependent on weather conditions and climate and requires high initial investment in order to construct heat exchange systems and integrate them with energy storages. Finally, several scientists are concerned about the effects of placing such structures in desert areas which are the natural habitat of wildlife. The vast number of mirrors can negative impact the animals that live in such areas and especially the endangered species on the verge of extinction.

Figure 23: A parabolic trough system

3.1.2 Wind Energy

Wind energy was one of the first sources of energy which was used by the human, as it was utilized for the movement of ships. It refers to the power generated by the wind and it is one of the most used renewable energy sources in todays' world with several applications. Its main use is to transmit mechanical energy to the wind turbines, which then will produce electricity. Through the years, the development of wind energy production technologies was rapid, which converted this type of energy source a viable and economically advantageous solution for energy production. Turbines' mechanical design improvements, their aerodynamic behavior and the use of better construction materials are some of the things that enhance and maximize the efficiency of those systems.

More specifically, wind turbines are engines which produce electrical energy from kinetic energy. A wind turbine consists of a column which is placed vertical to the ground and a turbine or a huge propeller at the top. There are small turbines which can be used to supply a house and there are massive turbines which are able to provide electrical energy to a whole power grid. Wind turbines can be clustered into two categories concerning their axis: vertical axis wind turbines and horizontal axis wind turbine. The latter, is the oldest category and the most used in wind farms.

Horizontal axis wind turbines are able to avoid extreme weather conditions such as heavy storms by adjusting their blades in order to protect them. These type of wind turbines have the ability to exploit the maximum amount of wind energy each time through their innovative design and they are made up of three main components: the nacelle, the tower and the rotor blades. In Figure 24 a farm of horizontal axis wind turbines is presented, while in Figure 25 the main parts of those constructions are shown. The tower constitutes the base where the nacelle and the rotor blades are placed. It is designed to keep the blades away from the ground with an average height between 25 and 100 meters and it is made of extremely strong variants of metal such as steel lattice and tabular steel. The higher turbines are placed the more efficient they are and the more electricity the tower can generate.

Concerning the cost and the performance of a horizontal wind turbine, the rotor blades or the rotating part of the turbine is one of the most crucial parts in its design. They are constructed from synthetic materials so as to be light and strong and their length and size are strictly connected with the height of the tower. The shape of blades is anatomical and focuses on maximizing their performance while inside them there are several electrical tools which control their behavior. Most of the horizontal wind turbines consist of three rotor blades. Finally, the nacelle is a solid tube made by fiberglass where all the main components of the wind turbine are placed. Two shafts (the main and the highspeed), a generator and a gearbox are some of the most important parts inside a nacelle which are synchronized and cooperate in order to produce low voltage electricity by exploiting the wind and the blades movement. The nacelle is the heaviest part of the wind turbine after the tower and it is placed on the top of it.

Figure 24: Horizontal wind turbines farm

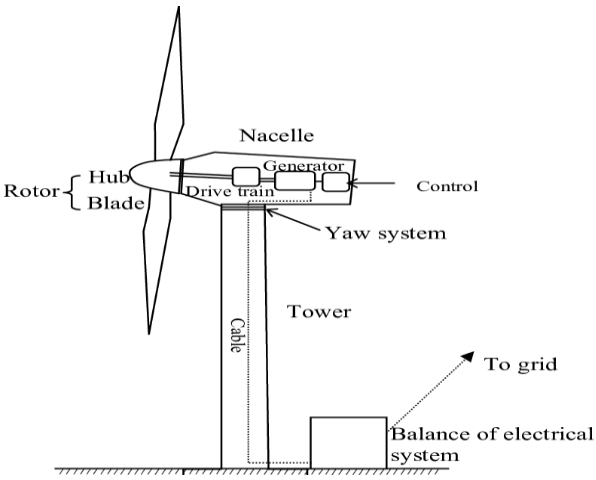

Figure 25: Horizontal wind turbine main parts

In case of vertical wind turbines, the rotor blades are placed perpendicular to the ground as can be shown in the farm of vertical axis wind turbines presented in Figure 26. This design allows the generator to be productive regardless the wind direction as the blades can rotate 360 degrees, which is the main advantage of this type of wind turbines. Thus, they are useful and more effective in areas with changing winds as they do not have to stop producing every time the wind direction changes. Additionally, the generator of these units is mounted close to the ground, which makes them more accessible for maintenance which constitutes another advantage when compared with the horizontal wind turbines. On the other hand, this type of installations definitely produces less energy due to the low height of the turbines and the vertical direction of the blades.

Figure 26: Vertical axis offshore wind farm

It worth to be mentioned that a wind farm or wind station can be defined as an area with several wind turbines which produce massive amounts of energy by converting the kinetic energy into electricity. These industrial parks do not contain only wind turbines but also meteorological masts, power cables and transformation stations as well. They can be placed in both land and sea with the latter having an extremely increased installation cost as their base is built at the bottom of the sea. Air density, wind speed and design of the turbine or even of wind farm are the main three factors which drastically effect the efficiency of these units.

Wind energy and wind turbines seems to be a viable and environmentally friendly solution for energy production. Wind is a natural phenomenon which is infinite, free and can be found almost everywhere around the world. Wind turbines and farms do not pollute the sea and air as they do not use harmful procedures and materials such as burning coal and do not dispose toxic chemicals into the environment. Except of these, wind farms do not abuse the land as the landscapes around them can be exploited for other purposes. However, their installation interferes with nature and the environment and the massive amount of land needs can negatively affect the landscapes. Additionally, the noise that they produce can disturb the harmony in the natural habitats around them, while thousands of birds are killed daily by the blades' movement. Last but not least, they lack adequate energy storage which is undoubtedly a vital weak point of wind power turbines and wind farms compared with conventional energy production units.

3.1.3 Hydropower Energy

Hydropower is considered to be probably the main source of renewable energy for electricity production [21]. It can be defined as the energy that comes from the water movement, or the energy which derives from a large amount of water falling from a certain altitude continuously. In other words, falling water contains kinetical energy which is converted into mechanical energy by a turbine and the latter is converted into electrical energy with the use of a generator. Several ways have been invented in order to exploit the kinetic energy of the streaming water, with the oldest one being the watermills and the most commonly and widely used nowadays being the hydroelectric power plants. The efficiency of the latter can vary from a few watts to several gigawatts and they can be divided in three main categories, concerning the water flow and the generating method: the pumped-storage plants, the conventional storage plants and the run-of-the-river plants.

The conventional storage hydropower plants use dams so as to store water in a reservoir. More specifically, they exploit the water's movement due to the height difference between the channels from which the water inserts the dam and those from which exits it. Dams are designed so as to keep only the necessary water in the reservoir and as this water falls, it moves a turbine in order to turn on a generator. The volume of flowing water and the height difference between the turbine and reservoir are the main determinant factors concerning the energy productivity of a dam. The more the water in the reservoir, the more the energy will be produced from the generator. Areas with massive rainfalls and rich springs are always preferred for these constructions for the aforementioned reasons. A significant advantage of this type of hydropower plants is that they have enough storage capacity in order to operate even for a few weeks. In Figure 27 the anatomy of a conventional storage hydropower plant is presented.

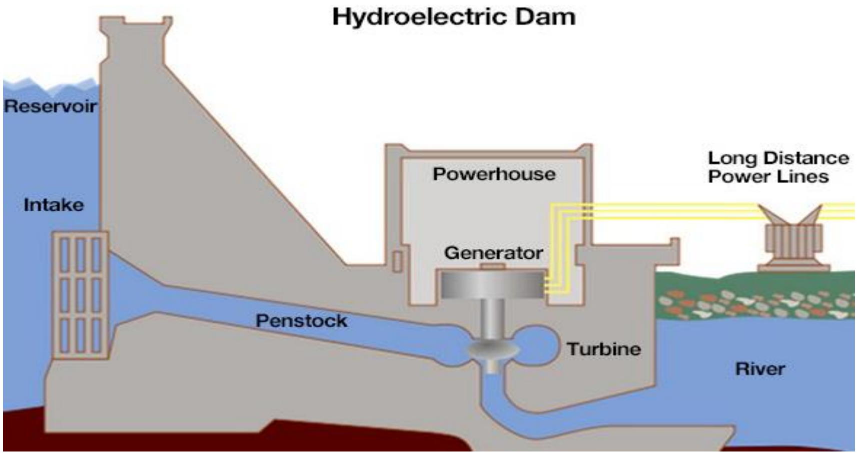

Figure 27: Conventional hydropower dam

The run-of-the-river plants, as their name suggests, are mainly installed in river or lakes, where natural reservoirs can be exploited. They take advantage of the natural water flow in a river so as to empower a turbine through a canal which then produces the energy. The advantage of this method is that provides the only water available for energy generation and the excess of water is unused. An example of such installation is given in Figure 28.

Figure 28: Run-of-the-river hydropower dam

Finally, the pumped storage hydropower plants operate in the same way that the conventional ones do by using the height difference and the water movement in reservoirs so as to produce electricity through generators. The key difference between the two types is that in the case of pumped storage, when there is low energy demand and water excess, this excess is used to pump water into reservoirs of higher-elevation until the demand is higher and the water returns to a lower reservoir. Due to the pumping process, the energy losses are extremely high which constitutes the main weakness of this type of hydropower plants. However, pumped storage installations are significantly useful when energy storage demand is high as they provide a high-profile energy storage system

Hydropower energy, as the aforementioned types of green energy, constitutes a clean energy source which is based on another infinite and natural good, the water and its movement. Water that is accumulated in reservoirs can also be used for other purposes such as irrigation, water supply and recreation. However, big-scale hydropower plants have received harsh critiques for their negative impact on the environment and in many cases, they are not considered as a green energy source. As the previous types of green energy plants, hydropower energy plants have extremely high cost of dams' and equipment and their installation can not only highly affect the natural habitat of many plants and animals, but also can lead to movement of populations and climate changes in the areas where located.

3.1.4 Biomass Energy

Biomass energy or bioenergy, is energy produced by living or once living organisms. In other words, it is a renewable organic material which derives from plants and animals. It constitutes an extremely important energy source and fuel for many developing countries, especially for cooking and heating. Biomass materials contain stored chemical energy from the sun which has occurred from the process of photosynthesis. Biomass sources for energy production vary and have several forms: wood and wood processing wastes, biogenic materials in municipal solid waste, animal manure and human sewage, biogenic materials in municipal solid waste, agricultural crops and waste material etc.

The aforementioned materials can be used with several ways for a variety of applications. The use of biomass materials can be summarized in the following five categories: feedstock, fuel, food, fiber and fertilized. Some of these materials are used as bioenergy fuels such as livestock waste and food waste, while others can be used to produce renewable natural gas or biogas such as the biogenic materials. At the same time, wood and other crops can be burned to produce heat, one of the simplest examples of the use of biomass fuel. Biofuel, is a fuel which derives after a certain preprocessing of some biomass materials. This fuel can be stored, transported and used whenever it is needed which makes biofuels a unique and quite useful fuel as the energy-on-demand option seems to be the most viable and efficient for the market. In summary, biomass can be transformed to energy through several procedures such as:

- Direct burning → heat
- Chemical conversion → liquid fuels, solid, gaseous
- Biological conversion → liquid fuels
- Thermochemical conversion → liquid fuels, gaseous

Biomass energy use has several advantages which are mainly economical. They are originated mainly from waste and other useless products which makes them extremely cheaper compared to fossil fuels and at the same time an efficient way to reduce the total waste. Additionally, it is a flexible energy source as can be used to create different products from different forms of organic materials. Just like sun, water and wind, bioenergy is widely available and renewable which provides distributed access to energy and reliability. As it is a natural part of the carbon cycle it doesn't burden the environment with excess carbon production like fossil-based sources and thus it can be characterized a green source of energy. On the other hand, uncontrolled biomass production can lead to deforestation with significantly negative impact on the environment. Also, it seems to be less efficient than fossil fuels and at the same time high qualities of fertilizers and pesticides are needed which results to an increasing air, solid and water pollution.

3.1.5 Marine Energy

Marine energy, or ocean energy, is the energy produced by the various forms of energy that the ocean water contains. It includes the energy produced by the kinetic energy of the oceanic waves, the tidal currents, the oceanic temperature differences and even the salinity of the water. Marine energy is currently used for energy production mainly through the forms of ocean thermal energy, wave energy and mainly tidal energy. These and several others form of ocean energy are still under research and development as the oceans are vast, unexplored and sometimes inaccessible aeras but they seem to have a great potential in world's need to replace conventional energy sources with the green ones.

Tidal energy is the most used marine energy form to generate energy. It comes from the tidal range, that derives from the sun's and moon's gravity which force the water to increase and decrease its level. When the water level is high, the water gets stored and when it drops, then the water activates a turbine which produce electricity. The greater the tidal variation is, the better the potential for energy production from this form of marine energy. In Figure 29 a group of tidal power plants is presented.

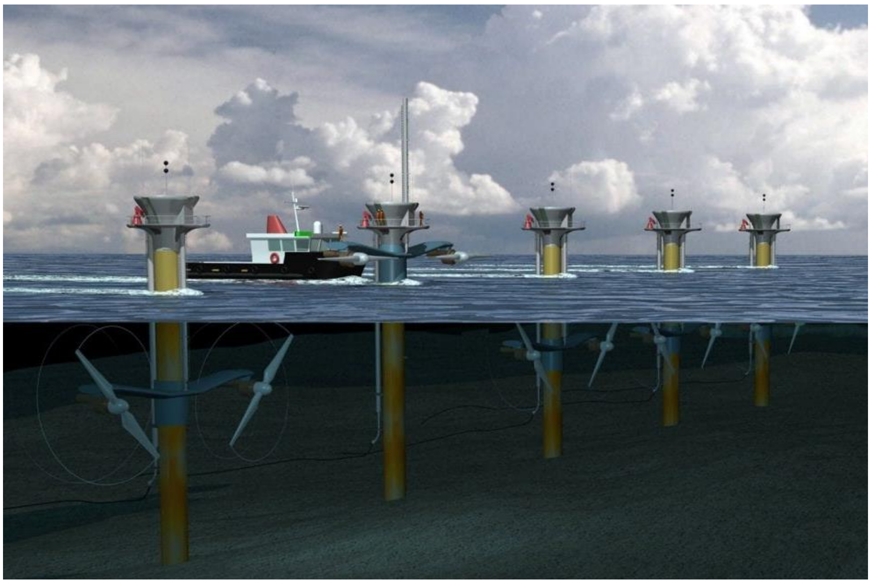

Figure 29: Tidal power plants

Wave energy utilizes the kinetic energy that waves contains and which is derived and transferred to the waves through the strong winds when they occur. This form of energy can be exploited by units that are installed near the shore or offshore and which consist of converters that can convert waves' kinetical energy to more beneficial types. Obviously, the larger the waves, the higher the energy produced by the converters.

Ocean thermal energy harnesses the thermal gradients or simply the temperature differences between deep ocean waters and ocean surface waters. Commonly, surface water seems to be warmer that water in greater depths. This difference in temperatures can be used by an engine which generates electricity from ocean heat. Larger differences of temperatures can lead to larger amounts of energy produced.

Marine energy has several advantages and great potential for a sustainable future. Like the aforementioned green sources, ocean energy does not emit greenhouse gasses when generated and it is renewable and widely available. It constitutes an ideal energy solution for islands and coastal and isolated areas, which most of the times lack in different forms of energy resources. On the other hand, this type of energy source requires units to be installed in extremely difficult locations and with significantly high equipment costs. Except of the costs and the installation difficulty, there are environmental issues which makes the use of marine energy plants controversial.

3.1.6 Geothermal Energy

Geothermal energy, as its name suggests is the energy produced by utilizing the natural heat of the Earth's surface deriving from its interiors. Depending on the temperature level, this type of energy can be used for several applications, from heating a household through a heat pump which extract the energy from the Earth, to supplying a whole city with electricity generated by a geothermal power plant. It is considered as a renewable energy source, as it is inexhaustible and at the same time the processes to utilize it has only small amount of consequence to the nature, while it helps to reduce the emissions of greenhouse gasses.

Geothermal energy can occur in different temperatures and depths but high temperature resources, which are the most valuable are usually found close to active volcanoes. The thermal energy of the earth is mainly stored in stone and water in the form of liquid or steam. Electric power generation and non-electric use are the two main methods of geothermal energy utilization and they are dependent to the type of the resource. The three basic categories of geothermal energy sources related with stone temperature exploitation: low temperature heat sources, hot aquifers of hydrothermal heat sources and high temperature heat or hot dry rocks. The first consists of sources with temperatures lower than 60 degrees Celsius and used on small scaling tasks, for agriculture tasks and air conditioning. Hot aquifers have temperatures between 80 and 250 degrees Celsius and can be utilized for low-efficiency electricity production or heating systems. The last category is mainly consisted of volcanic sources which have temperatures more than 250 degrees Celsius and can be used for electricity production of larger scale.

Figure 30: Geothermal energy power plant.

One of the several advantages of geothermal energy use is that is has close to zero emissions as it said before, which makes it unique even between green energy sources. Thus, it helps to reduce the world's energy dependence to fossil fuels and the deterioration of greenhouse effect. Additionally, it seems that it has huge potential as there are new technologies which being created to improve the energy process. There are an increasing number of projects to improve and grow this area of industry. With this rapid evolution many of the current cons of geothermal energy will be mitigated against. Concerning that cons, location restrictions, high initial and development costs and some environmental side effects which derives mainly through the emissions of gases during digging, are the most discussed and worrying.

3.1.7 Nuclear Energy

Nuclear energy or energy power is the energy contained in the core or nucleus of an atom. When bonds between protons and electrons within the atom core are broken, nuclear fission happens, a process through which a significant amount of energy is released. Nuclear power uses a radioactive fuel which is not constantly replenished but it does not release greenhouse gases when used for energy production and thus its environmental friendliness constitutes a controversial topic. It is worth mentioning that nuclear energy, is the second-largest source of low-carbon electricity in the world behind hydropower and its use seems to be essential for many countries in order to reach their energy needs without realizing greenhouse gases.

Concerning energy production, nuclear energy is mainly used to produce electricity. Uranium is the most common and suitable fuel used in nuclear energy plants for nuclear fission and can be used to power both domestic buildings and industries. As said before, it is a carbon-free electricity source which in parallel has high density as a nuclear fission can produce multiple times the amount of energy produced from burning an atom of fossil fuel. Also, it has a low production cost compared to fossil fuels and in contrast with the aforementioned green energy sources, nuclear power plants can be constructed almost everywhere without having a drastically negative impact on the environment around them. In Figure 30 a nuclear power plant is presented.

Figure 31: Nuclear power plant

However, there are several disadvantages concerning the nuclear power for energy production which have led to its exclusion from the official list of green energy sources. Most notably, the byproducts of nuclear power are radioactive waste and radioactive material which are extremely dangerous and cannot be destroyed. Actually, they must be transported and stored safely for long-term periods until these materials stop being dangerous for human health and for the environment in general. These periods could last even tens of thousands of years and thus the safety measures that should be taken must be extremely extensive. Although, the possibility of an accident is always there and radioactive disasters as history shows could be very destructive.

3.2 Climate Change and Renewable Energy Market

The reckless usage of conventional energy sources through the years has led humanity and Earth's viability to a crucial point. The need for change in the way energy is produced and used and the exploitation of other sustainable and green sources is more urgent than ever. In this chapter, the impact on the environment and the efforts of the European Union to set and impose a "green" framework which should be applied by the member states is presented. Finally, the energy market and the application of this framework in Greece is analyzed.

3.2.1 Climate Change

Through the years, the utilization of energy has aided the survival of humans and has improved their standard of living -mainly in the large urban areas. The industrial revolution led to the formation of capitalist economies and massive industries which are based on the daily production and consumption of vast amounts of goods and services. Despite the significant improvement of people's daily lives through the use of energy, massive productions of goods and services require enormous amounts of energy in order to be produced, which in turn requires unconsidered exploitation of natural resources and the negative impact of these processes on the environment is more obvious than ever.

The burning of such amounts of coal and petroleum in order to generate electricity produce fossil. Billions of machines all around the world contribute on this phenomenon, with industries, transportation (vehicles, ships, airplanes etc.), electricity production as said before and heating being the main causes of fossil fuels production and air pollution. Additionally, except of burning fossil fuels, the extensive use of fertilizers, the unconsidered destruction of landfills through waste decomposition in them and deforestation has led to the greenhouse effect which threatens the planet and its ecosystems. The Greenhouse effect is the way in which heat is trapped close to Earth's surface by greenhouse gases. These heat-trapping gases keep the planet warmer than it would be without them and due to this the average temperature on Earth tends to show a significant increase in recent years, known as climate change.

Climate change has several negative impacts on human's life and on ecosystems at all. WWF's (World Wide Fund for Nature) research has shown that about 33% of planet's life is in danger due to climate change. It is estimated that over one million flora and fauna species are threatened with extinction until 2050. Moreover, another harmful product of industrialization is the acid rain which derives when fossil fuels enter into the clouds and get mixed with rain or snow. Acid rain is extremely injurious for plants, lakes and rivers and certainly for the animals living in these ecosystems, as it contaminates and pollutes forests and the water.

Furthermore, climate change has led to the aggravation of extreme weather conditions which often lead to terrible natural disasters. Phenomena such as droughts, storms and floods become more and more intense, while temperature increase causes glaciers and sea ice melting and extreme heat waves. According to Intellectual Property Commission (IPCC), the last 100 years the Earth's average temperature has increased by 0.74 °C as can be shown in Figure 32. Finally, the toxic waste or simply the garbage of factories and power plants which are produced daily in massive amounts seems to be the biggest and maybe the unsolved problem of conventional energy production domain. Due to their production rate, toxic waste become unmanageable and they accumulate at the sea or at landfills, resulting in the destruction of habitats and in further emission of toxic gases.

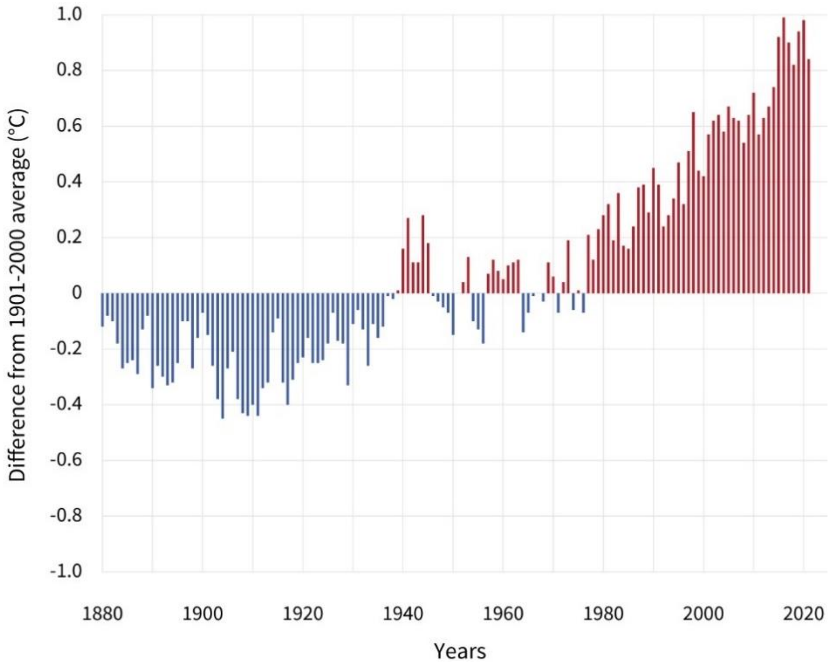

Figure 32: Global Average Surface Temperature

It is obvious that solutions for the aforementioned problems are necessary more than ever. A sustainable and viable future on Earth requires strong and decisive interventions in the way that energy is produced. Green energy or renewable energy came to the fore about 50 years ago (1970) due to the successive oil crises and now constitutes the only way to the sustainability of the planet and humanity. Generating energy that produces no greenhouse emissions from fossil fuels and reduces some types of air pollution is what green energy sources do. Additionally, it is worth mentioning that air pollution from fossil fuels cause health and economic costs which are about $8 billion per day globally. Today's challenge is to find ways to optimize and makes the renewable energy units more efficient in order to create sustainable economies and mainly a viable environment for every form of life on earth.

3.2.2 Global Mobilization & European Target Model

For the reasons that were presented in the previous sections, the importance of taking action and eliminating the factors which seems to enhance the greenhouse effect is greater than ever. If decisions had not been made at international level in previous years and a strict framework of rules and goals had not been defined, then the situation would probably have worsened much more and might would be already irreversible. In 1992, 154 countries (197 today) signed the United Nations Framework Convention on Climate Change (UNFCCC), which is a multilateral environmental agreement targeting to reduce the greenhouse emissions globally in order to avoid the destructive consequences of climate change and at the same time keep the ecosystems and the economies around the world stable. Created in 1988 by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO), Intergovernmental Panel on Climate Change (IPCC) has the objective to provide governments at all levels with scientific information that they can use to develop climate policies. IPCC often works per UNFCCC's request and provides regular assessments of the scientific basis of climate change, its impacts and future risks and options for adaptation and mitigation [22].

The first conference of the participating countries led to Kyoto Protocol which was the first official agreement with specific directions and targets to the parties on how the greenhouse emissions would be reduced [23]. Kyoto Protocol got into force in 2005 and its main principle was that developed countries should be separated from the developing ones as the industries of the first have extremely greater impact to the enhanced greenhouse effect and the climate change. For this reason, specific emissions targets were set only to developed countries, which had to reduce their greenhouse gases emission rates compared to 1990.

Unfortunately, Kyoto Protocol turned out to be incomplete and insufficient to address climate change in long-term scope due to the differentiation between developed and developing countries. Because developing countries lacked any commitments, the achievement of the goals by developed countries was proven not enough as the greenhouse emissions from developing countries was multiple times bigger. This happened because after the Kyoto Protocol agreement, many developed countries migrated their heavy industries to developing countries so as to avoid the penalties.

It was obvious that the only way to address climate change and take actions in an international level was the coordination and collaboration of all countries with no exceptions concerning their developmental level. The Paris agreement was the new improved international agreement which was signed in 2015 to replace the Kyoto Protocol and it has strict requirements from all member countries to develop national strategies in order to reduce greenhouse gas emissions. Additionally, a framework has been designed to assist countries within which financial support for the necessary infrastructure, climate technologies and the know-how for the incorporation of the previous in a country's industry and economy system are included. In this context, Nationally Determined Contributions (NDCs) are utilized. NDCs are outlines about the actions and the measures that each party should take in order to achieve its goals and are submitted every 5 years for inspection.

3.2.3 European Grean Deal and European Target Model

Concerning the application of Paris agreement in Europe, the European Union committed to implement the European Green Deal which is a long-term climate policy based on the main principles of Paris agreement. Its goal until 2050 is to transform Europe's economy to be resource efficient, environmentally sustainable and competitive, making Europe the first climate neutral continent. European Green Deal focuses on protecting communities and people who will be financially impacted by the transformation to green energy and sets specific objectives with certain solutions which should be made.

After achieving the previous objectives overall, the latest goals set by the European commission in 2020 is to reduce at least 55% the greenhouse gas emissions by 2030, increase the renewable energy share from 32% to 40% and show improvements in energy efficiency from 32.5% to at least 36% (always compared to 1990) [24]. Figure 33 shows the progress that EU members have shown in total, concerning the renewable energy use. The red point was the objective for 2020 which was achieved, the blue point is the current 2030 target and the orange point is the proposed 2030 target.

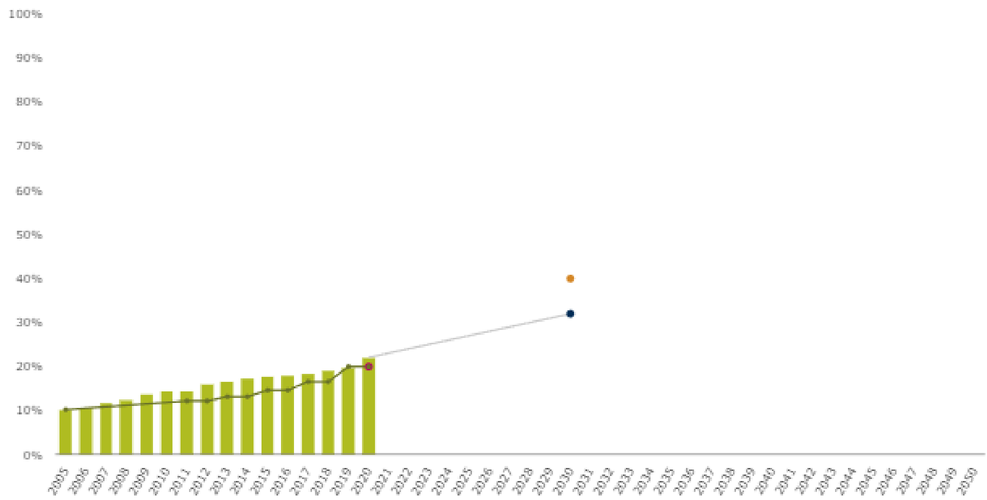

Figure 33: Progress towards renewable energy source targets for EU members

However, in order for this framework to be properly and fairly applied within the member states of EU, there were several issues mainly concerning the reliability between the parties and the coordination of the total energy market. It seems to be significantly important for Europe to have interconnected energy systems through which the energy can freely flow across the borders and address issues concerning the instability of the weather conditions which can lead to fluctuations on the amounts of the renewable energy produced in some places. Additionally, integrating different national markets with different principles and different architectures can be challenging and extremely complex.

In this context, European commission proposed a “Target Model” for electricity markets as the basis for the development of a single energy market in Europe. The EU Target Model enhances harmonization in the European electricity market as an integrated market by establishing common rules and directions for the proper use of cross-border capacity and electricity movement. This model achieves:

- optimization in the use of the transmission system capacity through synchronization of system operators
- reliable prices and liquidity in the allocation of the node capacity for the day-ahead market
- the efficient operation of the forward markets and the efficient design of the intra-day markets for the allocation of the node capacity difference caused by imbalance settlement between the day ahead scheduling (DAS) prediction and the actual demand

The European Union’s plan to achieve a single and integrated energy market, consider that the common market should be consist of the following four individual different markets:

- The Day-ahead Market, which concerns transactions for the purchase and sale electricity with the obligation of physical delivery on the next day. It is worth mentioning, that there should be a balance in energy when the markets close every day, which means that planned production must be equal or close to next day’s demand. Through this market, the exposure of price volatility is reduced and the reliability to the system is strengthen. Energy producers are forced to participate in Day-ahead Market.

- The Intraday Market, which concerns transactions for the purchase and sale electricity with the obligation of physical delivery on the same day. The participation to them is not obligatory and they give the participants the opportunity to be aware about possible deviations in production and demand. This happens as this market is made after the previous day's day-ahead market closure.
- The Forward Market are useful as they ensure transparency in trading and protect participants from price volatility. Future Contracts and agreements for predetermining prices and quality are used in order for this safety framework to be accomplished.
- The Balancing Market consist of the phases of balancing capacity market, balancing energy market and imbalances settlement and is a fundamental process in EU's electricity target model as it ensures the balance between supply and demand.

3.2.4 Greek Energy Market

The energy market in Greece changed drastically in 1950 with the establishment of the Public Power Corporation (PPC) which was created in order to improve the previous model of energy production, pricing and distribution used in Greece. PPC achieved to provide electricity everywhere and at the same time to apply the same pricing model without discriminations between remote and urban areas by expanding the network and consolidating the local energy companies. More simply, through PPC, a unified energy policy was applied throughout the country. Additionally, a few years later (in 1999), in order for Greece to harmonize with the European rules that had been set regarding the internal energy market, the Regulatory Authority for Energy (RAE) was founded. RAE is an independent regulatory authority which regulates the energy market in Greece and monitor the proper application of consumer protection measures. In addition, it works as a controller for Independent System Operators.

RAE also regulates Hellenic Electricity Distribution Network Operator (HEDNO), which is a PPC's subsidiary in charge of operating, managing and developing the Greek electricity distribution network. Furthermore, in 2011, IPTO or Independent Power Transmission Operator S.A was founded, which is responsible for the operation, control, maintenance and development of the Hellenic Electricity Transmission System. It ensures the country's supply with electricity in an adequate, safe, efficient and reliable manner, as well as the operation of the electricity market for transactions outside the Day Ahead Scheduling, always compatible with the principles of transparency, equality and free competition. IPTO's compliance with the requirements applicable to the Independent Transmission Operator model was certified by the RAE in December 2012. Concerning the energy market share,

Table 1 shows the top seven energy suppliers in Greek territory, based on February 2022 data [25].

Table 1: Top 7 energy suppliers in Greece, February 2022

Supplier	**Share**
PPC	64.59%
Mytilineos	6.97%
Heron	6.48%
Elpedison	5.90%
NRG	4.36%
Watt and Volt	2.33%
Volterra	2.09%

Regarding the application of the European Target model in Greece, it started to operate on a transitional basis from November 2020 and was supposed to be fully implemented in the first months of 2022. The operation of forward, day-ahead and intra-day markets has been assigned to the Hellenic Energy Exchange S.A. (HEnEx), while IPTO is responsible for the operation of the balancing market. Concerning the renewable energy market and the goals that have been set for 2030 by the EU, RES units must now have an equal role in the electricity system and equal terms of participation in the electricity market and for this reason, the institutional framework is adjusted to this specific perspective. The FoSE (Cumulative Representation Bodies in Greek) are the new entities of the electricity market through which the full integration of RES units into the electricity market will take place.

FoSEs are specialized innovative companies that use modern tools and are asked to optimize the commercial management of green energy units by minimizing balancing costs with accurate forecasting of their production. Therefore, in the coming years, when participation of renewable energy sources in the market will be done on competitive market terms, FoSEs or otherwise green aggregators will become more and more necessary for the management of many RES projects. According to Law 4414/2016, RES projects that have a capacity of more than 3 MW for wind and more than 500 kw for photovoltaics undertake the obligation to participate in the wholesale market and at the same time undertake prices balancing obligations. This means that if they declare more or less power available than the actual one, they will bear the costs resulting from the deviation.

The application of the target model transfers the responsibility of balancing the market from the IPTO to the RES units, while the commercial management of the green units is done by the FoSEs. FoSEs limit the deviation between forecast and actual production and consequently the variability of the production of RES plants. This variability and uncertainty will be further reduced in the future when renewable energy storage systems are promoted and implemented in the market. Concerning 0ctober 2020 data [26], FoSEs' market share is shown in Table 2:

Table 2: FoSEs' market share, October 2020

FoSE supplier	**Share**
OPTIMUS ENERGY	46.5%
ELPEDISON	17,2%
MYTILINEOS	15,8%
RENOPTIPOWER	8,3%
INACCESS	5,3%
FOSETEK	2,1%
MOH	1,8%
FORENA	1,0%
EUNICE TRADING	0,9%
SOLARENERGY	0,4%
WOOTIS	0,3%
VIOLAR	0,3%

In general, Greece nearly doubled its share from renewable energy sources between 2004 and 2017 (from 6.9% to 15.5%) and aims to reach the EU goal of 2030 of 27% [27]. National energy and climate plan (NECP) which was signed by the Greek state in order for Greece to meet the EU's energy and climate targets for 2030 includes the following goals [28]:

- RES share in gross final energy consumption – at least 35%
- RES share in final consumption for heating and cooling – at least 43%
- RES share in gross electricity consumption – at least 61%
- RES share in final consumption for transport – at least 19%

Both in Greece and in Europe, the efficient utilization of green energy units is more urgent than ever. The target model which is established and used in European Union, requires increased energy production from renewable sources, which in turn requires efficient forecasting. The application of sufficient forecasting methods on green energy production will enhance not only the energy producers' position concerning the financial offers that propose to the market, but also will significantly contribute to the energy balance in the market. Incomplete renewable energy production forecasting, will lead to mandatory coverage of the energy needs from conventional energy sources.

3.3 Forecasting for Renewable Energy Production

As shown in the previous sections, the importance of renewable energy exploitation is more than obvious and for this reason, the construction of renewable energy units has grown exponentially the recent years. At the same time, the need of maximum proper utilization of these units has led the scientific community to the extensive development and improvement of several machine learning forecasting methods for renewable energy production. In this section, the challenges of applying forecasting methods on renewable energy production are presented, alongside with the benefits of this process. Moreover, the way that this kind of data are modeled and exploited for predictions are analyzed, focusing on machine learning approaches which is the topic of this thesis.

3.3.1 Challenges and Benefits of Renewable Energy Forecasting

The use of renewable energy sources seems to be a key point for a sustainable future on Earth and the forecasting concerning the energy production from these types of sources has become a topic which attracts the attention of the scientific community all around the world. Efficient forecasting can offer extremely important advantages regarding not only the environment sustainability but also for ensuring the energy market reliability. At the same time, there are several challenges concerning unstable factors that can influence the production of green energy, which should be taken into consideration.

In the energy market sector, energy production should always meet the energy demand. This balance between supply and demand is the most crucial part of the market as it is the only way to ensure the proper and the sustainable operation of power grids and their systems and at the same time the uninterrupted and fair coverage of the market's energy needs. Suppliers and producers need forecasting tools in order to schedule the amounts of the energy which will be produced and to determine if these amounts will be enough so as to cover the market's demand. If not, conventional energy sources will be used to complete the gap between demand and supply, leading to the aforementioned negative impacts on the environment and humanity. Moreover, electricity cannot be stored in massive quantities as most of the times should be transformed in other forms of energy which is significantly costly and space demanding. For this reason, with precise forecasting in green energy production, the needs for extreme energy storage demands can be eliminated.

Additionally, as the electricity industry becomes more and more competitive with thousands of suppliers, producers and buyers taking part in the market, the efficient forecasting for accurate energy production can facilitate healthy competition by decreasing the risks for producers and allowing them to make more competitive offers and more accurate contracts. Furthermore, it is vital to keep the energy prices stable over time and to avoid extreme values which can lead to financial and energy crisis. The energy prices are closely related and extremely sensitive to any changes between the production and consumption balance and the efficient forecasting will increase the reliability and the stability of the energy market and economies in general.

However, there are several challenges concerning the renewable energy production which should be taken into consideration. The performance of green energy plants, especially in the cases of wind and solar energy, is highly dependent on the weather conditions which are continuously changing and varying from country to country or ever from area to area. Additionally, except of geographic diversity, the green power plants' performance is also dependent on the quality of the installations' materials and mainly on the complementing technologies that are used in each grid.

Due to the previous factors, the renewable energy sources exhibit extremely high sporadicity, variability and randomness and accurate forecasting methods seems to be vital in order to strengthen the reliability and the stability of the green power grids [29]. For this reason, the efforts for exploitation and optimization of new or even existing forecasting methods becomes more urgent. Machine learning approaches compared to statistical ones seem to have greater potential as they are able to parametrize and take into consideration as many variables as needed. In this way, they are able to offer more reliable and justified results and most of the times even faster.

3.3.2 Machine Learning Approaches

Machine learning algorithms and applications can offer great improvements to renewable energy forecasting in terms of its production. Since green energy production relies on nature, the capacity and the efficiency of renewable energy plants can vary. In this thesis, datasets concerning solar and power energy plants will be exploited which are the two most related with weather conditions green energy sources. Different weather conditions such as wind speed, cloud appearance or sun heat intensity can lead to varying amounts of power produced and machine learning tools can enhance the weather forecasting process and thus the prediction of the energy production.

Weather conditions can vary between countries or even cities and deviate between different periods of the year or even different hours of the day. These and other important data closely related with the efficiency of green energy power plants can be modeled and exploited with various machine learning models, most of the times in the form of timeseries data, in order to lead in comprehensive, reliable and fast predictions. Until today, several approaches and frameworks using machine learning algorithms have been developed, concerning the exploitation and optimization of wind and solar energy plants, leading to interesting results.

When it comes to solar energy, in [30] a method using deep recurrent neural networks (DRNNs) to predict the solar irradiance is presented. The method was tested on real data from Canada and showed that it outperforms several common methods. Additionally, the authors of [31] developed a simple framework for solar irradiance forecasting. This framework focuses on the optimization of KNN and ANN algorithms for a specific range of time horizons and it seems to outperform simpler forecasting models.

In [32], three different deep learning architectures were used in order to learn the relationship between sky appearance and photovoltaic installations' future output. Each architecture was tested on images from Japan and on a dataset of photovoltaic power values, with LSTMs achieving the best RMSE score. Finally, in [33] a combination of state-of-the-art machine learning algorithms and a physical model was used so as to calculate the actual energy generation of a photovoltaic plant and minimizing at the same time the errors introduced by the weather forecasting tools which are utilized. The application of this engine in a photovoltaic installation in Greece showed great potential.

Regarding the wind energy, in [34] Jie Chen et al. proposed the EnsemLSTM, a novel method for predicting the wind speed. EnsemLSTM uses a nonlinear-learning ensemble of LSTMs, Support Vector Regression Machine (SVRM) and External Optimization algorithm (EO). It was tested on two datasets with data collected from a wind farm. Results showed that its forecasting outperforms other popular models such as ANNLSTM and MeanLSTM. Furthermore, in [35] a hybrid variational decomposition model (HVDM) is presented for energy production forecasting in hybrid PV-Wind farms and showed promising results.

Ramon Gomes da Silva et al. in [36], proposed a novel decomposition-ensemble learning approach which combines Complete Ensemble Empirical Mode Decomposition (CEEMD) and Stacking-ensemble learning (STACK) based on Machine Learning algorithms such as kNN and SVR to forecast the wind energy of a turbine in a wind farm in Brazil. Results indicate that the proposed models outperform the CEEMD and STACK in all forecasting horizons. Finally, in [37] GRUs and LSTMs were used in order to forecast a country-wide wind power production in Germany. Results revealed that GRUs are more accurate and at the same time can learn faster over long sequences compared to ARIM and SVR literature approaches.

In addition, there is a variety of timeseries forecasting models which were not developed specifically for predictions in renewable energy data but are more general and universally recognized as they have presented impressive results. In [38], Salinas et al., propose a novel methodology named DeepAR for probabilistic forecasting with high accuracy. DeepAR is based on an autoregressive RNN which is trained on a massive number of related timeseries data. This framework was evaluated on different real-world forecasting datasets and produced forecasts with higher accuracy than other state-of-the-art methods such as Snyder, ETS and ISSM. In 2020, the google cloud ai team published Temporal Fusion Transformer (TFT), an innovative attention-based architecture for multi-horizon forecasting [39]. Temporal Fusion Transformer (TFT) not only achieves high forecasting performance by using recurrent layers, but also offers interpretable insights into temporal dynamics with the use of self-attention layers for long-term dependencies. TFT was tested in several real-world datasets and showed great improvements concerning the accuracy compared to existing benchmarks such as DeepAR, ARIMA and ETS.

In [40] Boris N. Oreshkin et al. present N-BEATS, a deep neural network architecture for time series forecasting. N-BEATS constitutes an interpretable, fast to train and widely applicable with no modification solution which consists of a deep fully-connected layers network with backward and forward residual links. This architecture was evaluated on various famous datasets and some of its configurations showed state-of-the-art performances, with accuracy improvements even 11% higher compared with previous statistical benchmarks.

Finally, Oskar Triebe et al. from Facebook research team propose NeuralProphet in [41]. Facebook Prophet, which is the ancestor of Neural Prophet, is difficult to be extended due to its backend characteristics and at the same time lacks local context, an important part for short-term forecasting. Thus, NeuralProphet came to give solutions to these weaknesses as a scalable, explainable and user-friendly standard for forecasting frameworks. It constitutes a PyTorch-based hybrid forecasting framework which is configured with deep learning methods and can be extended easily. After being tested in several real-world datasets, the results showed that NeuralProphet outperforms its successor in terms of quality and interpretability.

Chapter 4

Solar Generation Forecasting with Trasfer Learning

In the previous chapter, the current state of renewable energy adoption in the market was evaluated and several actions were identified as necessary to increase renewable energy penetration. One key action is the development of reliable methods for forecasting the production from these sources. This can be difficult as weather conditions can greatly impact production and the intermittent nature of it leads to a non-reliable and difficult to predict and estimate production. However, accurate forecasting can provide operators with the necessary information to make informed decisions that allow for the most efficient use of generated renewable energy generated. High performance neural network architecture design and optimization is a challenging task which usually requires significant resources, often involving thousands of iterations and a large number of diverse trials and tests.

The main focus of this thesis is to design one or more pretrained models for solar power generation forecasting and explore their prediction performance for the next 24-hours ahead. The models are trained using a large dataset of energy production data collected from various sources across different European countries, and their purpose is to generate accurate forecasts for these countries and for countries that do not have sufficient data to train a new model. These models will eventually be compared with other baseline approaches which are trained on only one country's data.

This chapter concerns the implementation part of this dissertation -which focuses on solar energy generation and it is formulated as follows. Firstly, a thorough description of the data used is given, including a breakdown of the data preprocessing methods applied. Subsequently, the methodology followed is given and the experimental results are presented, with their significance being evaluated using predetermined evaluation metrics. Finally, the baseline models used and their respective results are analyzed.

4.1 Exploratory Data Analysis

Two datasets were exploited for the training and evaluation of the models, the first for weather data exploitation and the second one for energy data. Both datasets were acquired from Open Power System Data Platform [42], which offers a variety of energy sector related datasets for energy system models. The selected datasets are in the form of timeseries and concern renewable energy production. After the appropriate concatenation of the two datasets and the application of several preprocessing techniques which are discussed in the next sections, the final dataset was formed. The next two subsections give an overview and a summary analysis of each of the two datasets.

4.1.1 Energy Data

The second dataset is the one related with the energy information [43]. This package includes a variety of time-series data pertinent to power system modeling. This data encompasses electricity prices, consumption and the generation and availability of wind and solar power. The data is separated by country, control area or bidding zone and covers countries within the EU as well as some nearby ones. The information provided is at a granularity of one hour and covers a period of almost 5 years (01/01/2015 to 30/09/2020). This thesis exploits the solar generation data for a variety of European countries. In contrast with weather dataset, several missing values can be found in energy dataset. This derives mainly from the fact that this dataset contains data for certain time intervals and different countries did not start producing solar energy at the same time. The procedure followed to alleviate the problem of missing values is described in Section 4.2.2.

In Figure 34 and Figure 35 two indicative examples of the solar generation data are given. The figures present the solar energy generation from 2015 to 2020 in Germany and Greece respectively. As in the previous plots, seasonality pattern can be easily recognized in both plots. Also, it is clearly visible that Germany's solar generation has greater fluctuations compared to Greece's. That makes sense as solar generation is mainly related with the weather conditions and especially with the presence or the absence of the sun. Greece is a country with more sunshine through the whole year compared to Germany and for this reason solar energy generation in Greece seems to be more stable.

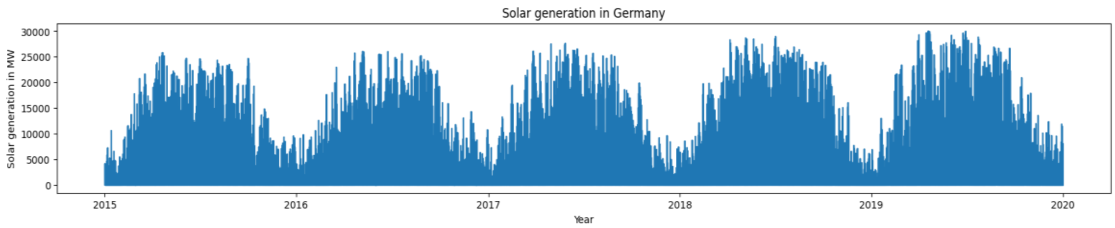

Figure 34: Hourly solar generation in Germany from 2015 to 2020

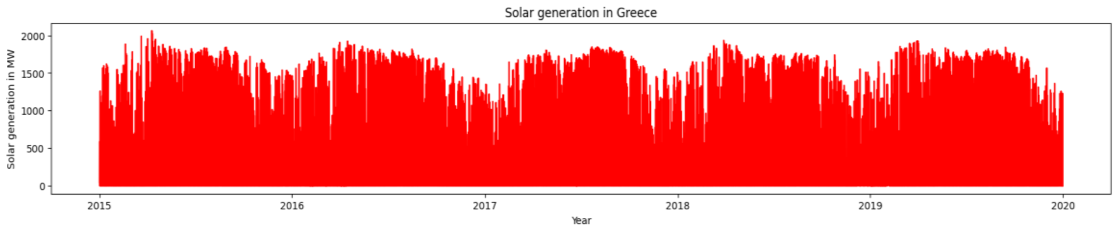

Figure 35: Hourly solar generation in Greece from 2015 to 2020

4.1.2 Weather Data

This data package [44] includes hourly measurements of radiation and temperature for 26 European countries for a period of 39 years (01/01/1980 to 31/12/2019). For the purposes of this thesis, only the measurements between 2015 and 2019 are taken into consideration, so as to span the same the same period as the energy data. Temperature measurements are on the Celsius scale, while the radiation measurements indicate the watts of radiation per square meter ($W/m^{2)}$). The great quality of this dataset is revealed from the fact that the is no missing values in any of its columns.

In Figure 36, Figure 37 and Figure 38 three indicative examples are given. In the first two plots, the weather conditions -concerning the temperature in both Germany and Greece are presented for the 2015-2019. The seasonal pattern can be easily recognized in both cases, as temperatures seems to reach their lowest value during winters and the highest value during summers. Seasonality is one of the main properties of timeseries as shown in Section 2.4.1. In Figure 38, Greece's solar radiation values are presented, for the period 2015-2019 and the seasonality component can be easily recognized again.

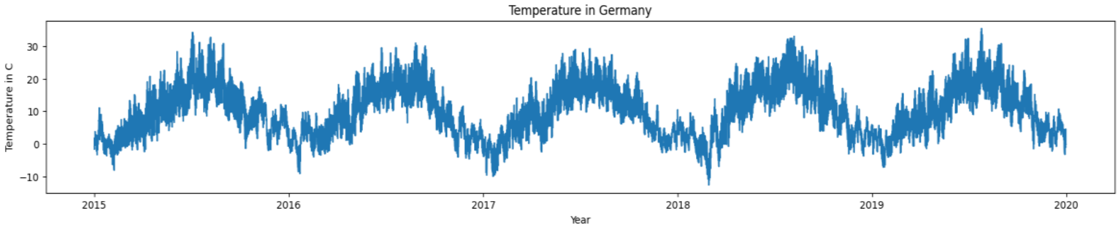

Figure 36: Hourly temperature in Germany from 2015 to 2020

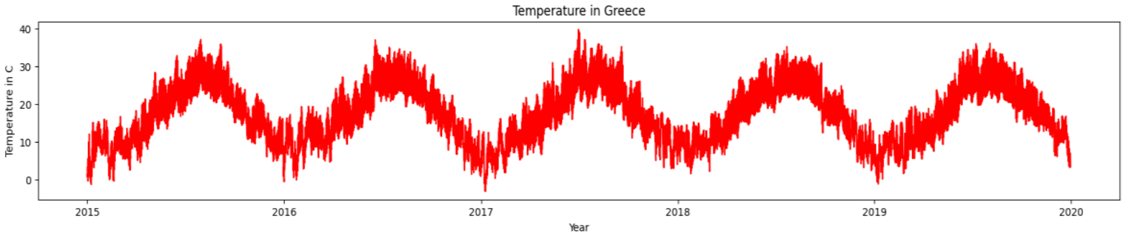

Figure 37: Hourly temperature in Greece from 2015 to 2020

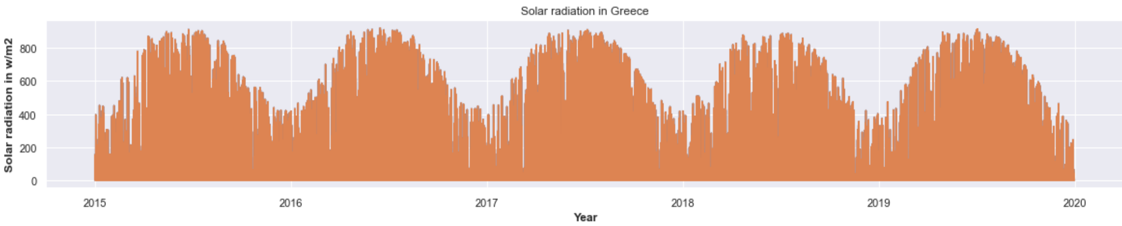

Figure 38: Hourly solar radiation in Greece from 2015 to 2020

4.2 Data Preprocessing

For the preparation of the dataset, a sequence of different techniques was applied, so as to transform the data into the propriate form for the prediction tasks that followed and the models that were being used. In the next subsections, details concerning the datasets' concatenation, the missing values' problem alleviation and the windowing method which were applied are given.

4.2.1 Datasets' Concatenation

As mentioned in the previous sections, for the preparation of this thesis, two datasets were exploited. The weather and energy datasets were presented in Section 3.5 and their concatenation constitutes the final dataset. Before datasets' concatenation, the useless columns from each dataset got removed (##_load_actual_entsoe_transparecy, ##_price_day_ahead, ##_wind_onshore_generation_actual, ##_radiation_diffuse_horizontal etc.). Afterwards, a mask was applied to each dataset so as to keep a specific time interval (2015-01-01 00:00:00 - 2019-12-31 19:00:00) which would be exactly the same for both datasets. Finally, one extra columns, month, was added to the dataset as an extra feature. This feature exploits the information which is given from the timestamp column and gives the specific month number that each example (measurement) took place, which could be a valuable insight. After datasets' combination, a dataset of almost 350.000 rows and 63 columns arose. As shown in Figure 39, the two main selected features (temperature, solar radiation) are highly related with the target variable (solar generation). The correlation matrix of this figure, concerns the data which are related with Belgium; however, the same correlations apply to the other countries too.

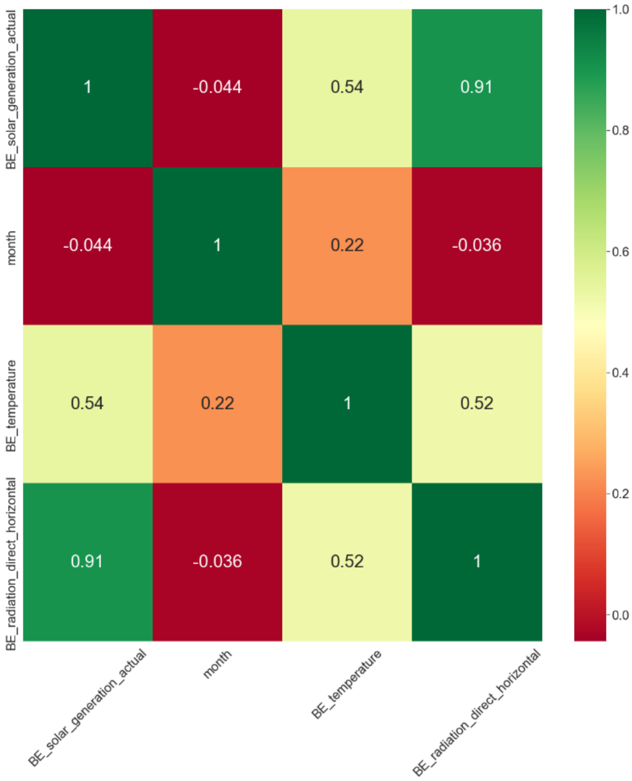

Figure 39: Features correlation matrix for Belgium

4.2.2 Handling Missing Values

Dealing with missing values in time series data poses a challenge as maintaining regular patterns and accurately estimating highs and lows is crucial. To address this, two techniques for imputing missing values were used, based on the number of consecutive missing values in each country column. For countries like Germany, which have a limited number of missing values, a simple imputation approach was used, using the nearby values. However, for countries which have extended stretches of missing values (e.g., more than 30 in a row), the imputation process involves the separation of the timeseries into parts. The first part encompasses all data up to the beginning of the period of missing values, while the second section covers data from the end of the missing values period onwards, until the next period of missing values (if exists), and so on and so forth.

Finally, there were some countries which showed an extremely high number of missing values concerning their solar generation measurements as shown in Figure 40. More specifically, solar data for Hungary, Croatia and Poland seem to consist of more than 90% of missing values and for this reason, these three countries were dropped from the dataset. After the imputation techniques, the dataset had been transformed in a list of lists. Each country is represented as a list which consists of one or more dataframes. These dataframes are the parts that arose from the application of the missing values splitting technique.

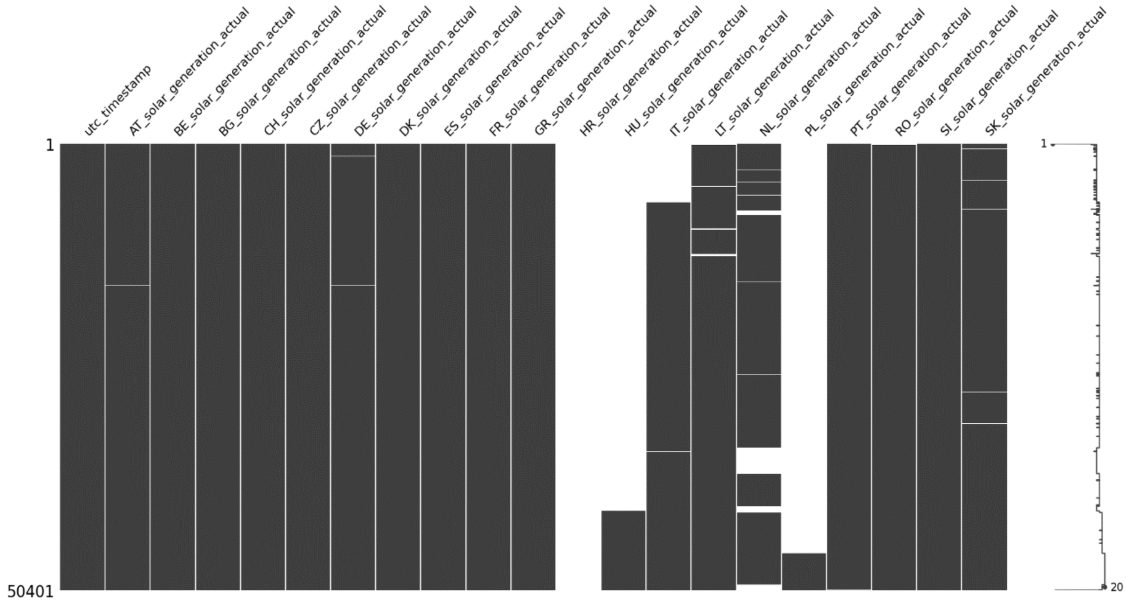

Figure 40: Solar generation missing values for each county

4.2.3 Countries Selection

After handling the missing values and dropping the countries which have more than 90% of missing values within their data, the remaining countries in the dataset were 20. Subsequently, due to limited computing resources as well as to save time during the multiple tests that follow, a further re-selection of countries was decided, through which the data set would shrink even more. The selection of the new subset of countries was done in such way as not to spoil the diversity of the data. The goal was to find countries which have the same or very similar distributions in all of their 3 features (temperature, solar generation, solar radiation) and then keep one of them. To achieve this, a normalization technique was first applied to the whole dataset, so that the data would be at the same scale (0-1) before the comparison took place. Afterwards, several histograms were created in order to observe the distribution of the values for each feature of the dataset.

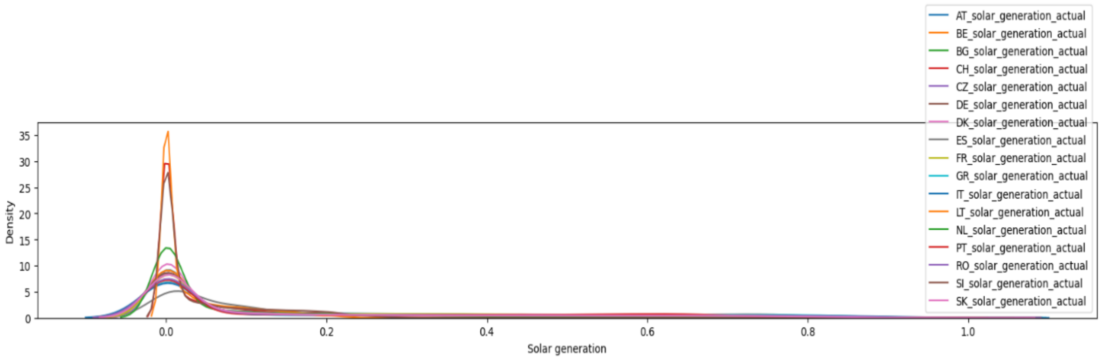

Figure 41: Solar generation values' distribution for each country

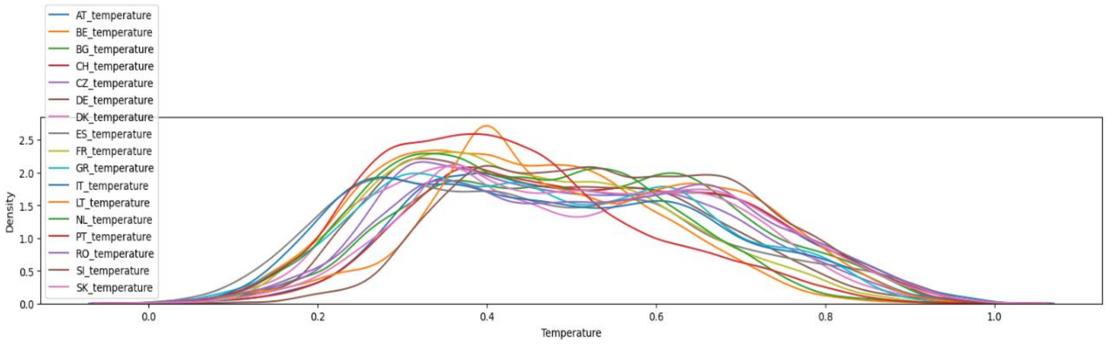

Figure 42: Temperature values' distribution for each country

In Figure 41 and Figure 42 the distributions of solar generation values per country and the temperature values distributions are presented. It is easily perceived that the distributions between some countries are totally different, while between some other countries seem to be very similar. To find out which distributions are more similar, the individual histograms should be observed. In the following figures an indicative example of how the comparison took place is given.

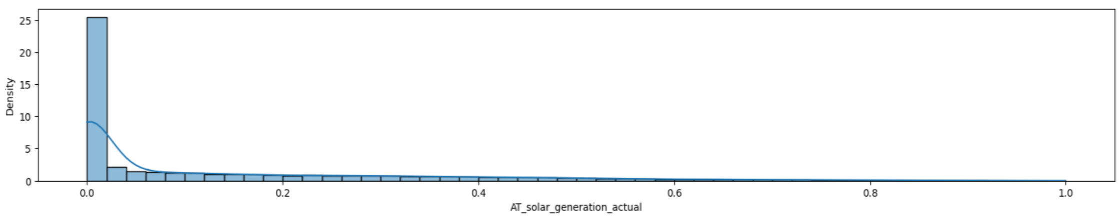

Figure 43: Austria's solar generation values' distribution

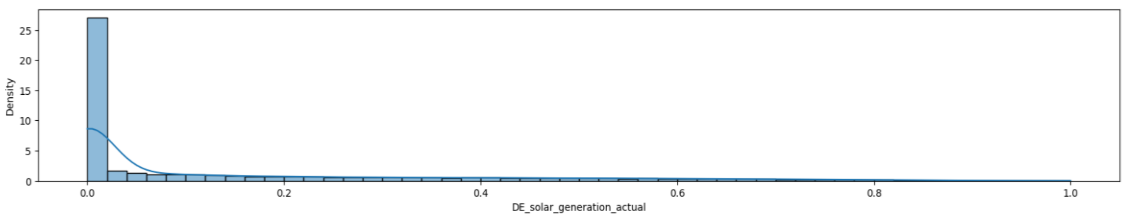

Figure 44: Germany's solar generation values' distribution

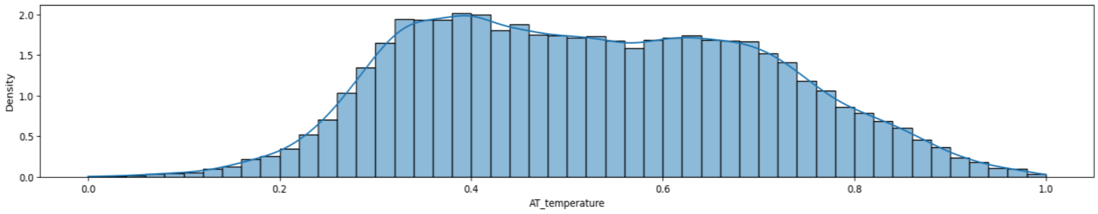

Figure 45: Austria's weather values' distribution

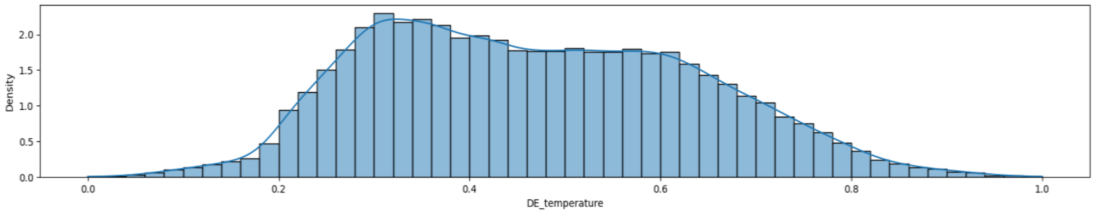

Figure 46: Germany's weather values' distribution

It can be easily seen in Figure 43 and Figure 44 that the distributions of solar generation values are almost the same between Austria and Germany in the scale of 0-1. Figure 45 and Figure 46 show that the same applies to the distributions of weather values between these two countries. Finally, before dropping one of the two countries, a comparison between their solar radiation distribution should take place.

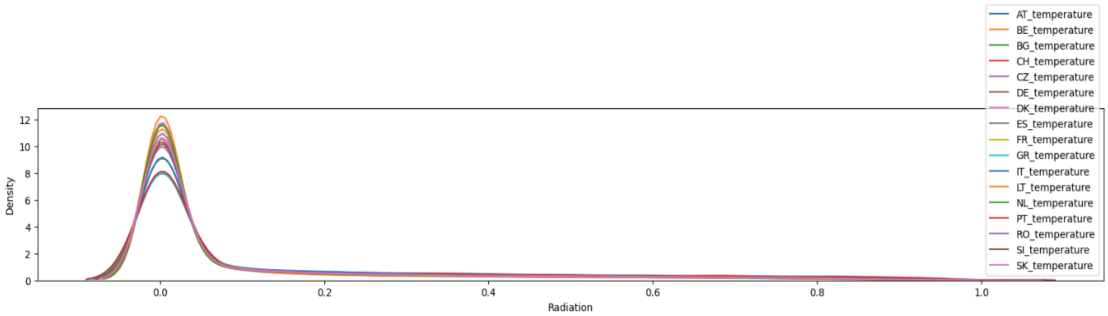

Figure 47: Solar radiation value's distribution for each country

Figure 47 reveals that almost all of the countries have very similar distribution when it comes to their solar radiation feature. For this reason, it was decided to select countries just by comparing the distributions of the previous two features (solar generation and temperature) as explained previously with Figures 43-46. From this comparison, five different groups of countries with similar distributions emerged: [Austria, Belgium, Czech Republic, Germany, Denmark], [Bulgaria, Spain, Greece, Italy, Portugal, Romania], [Switzerland, Lithuania, Slovenia] and [France, Netherlands, Slovakia]. One or two countries were selected from each group so as to end up to the final group of countries which are the following: Belgium, Denmark, France, Italy, Lithuania, Netherlands, Portugal and Slovakia.

Table 3: Final dataset example

utc_timestamp	BE_solar_generation_actual	BE_temperature	BE_radiation_direct_horizontal	month
2015-01-01 08:00:00	0.034217	-1.013	3.7333	1.0
2015-01-01 09:00:00	0.138437	0.187	59.1418	1.0
2015-01-01 10:00:00	0.273233	1.325	112.3332	1.0
...	...	...	...	...
2019-12-30 23:00:00	0.000000	-0.177	0.0000	12.0

In Table 3, the form of the final dataset is given. Finally, it is worth mentioning that the dataset was normalized before being used for models' training and predictions. The normalization technique which was applied to the whole dataset before, concerns only the comparison between the data's distributions and cannot be used in general. This arises from the fact that the huge differences which can be seen between the values of different countries, can lead to misleading or uninterpretable results in the prediction process. Data between different countries can be in totally different scales and must be normalized separately. For this reason, a scaling technique was applied to each country separately and each scaler was stored in a list. The scalers' list will be used again in the prediction step, where inversed scaling is applied to each country to give us the final result. The mean real value of target variable (solar_generation) for each country is given in Table 4.

Table 4: Solar generation mean real value for each country

Mean real value							
BE	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**
364	93	1029	2074	8	214	97	59

4.2.4 Windowing

The sliding window method or windowing is a technique where a small window of time is used as input data to train the model and then the model is tested in the next window which is used as the output, as described in Section 2.4.3. The forecast horizon is the set of points in the future that the model will predict, while the history window is the set of points in the past that will be used as the model's input. The windowing function, as shown in Figure 48, is responsible for slicing the dataset into proper windows and creating examples with the model's inputs and outputs.

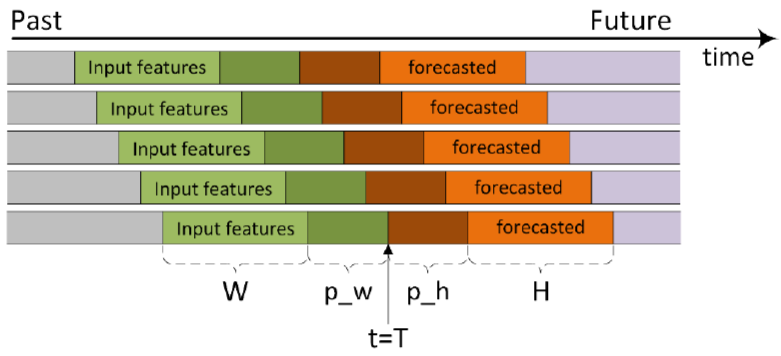

Figure 48: Windowing function

An indicative example of datasets resulting from a windowing method application are shown in Figure 48, where W=History Window, H=Horizon, p_w=history offset and p_h=future offset. In this thesis, the windowing function returns six lists: X_train_list, y_train_list, X_test_list, y_test_list, X_val_list, y_val_list which are the train, test and validation sets for both X and y that will be used by the models. Each list constitutes of several windows or simply of several lists. Table 5 shows approximately the amount of data that each set contains.

Table 5: Train – Validation -Test sets

	Train	**Validation**	**Test**
Percentage	60%	20%	20%
Hours	210.240	70.080	70.080
Days	8.760	2.920	2.920

4.3 Experiments Methodology and Results

This thesis evaluated several models, all trained on the same dataset with consistent history and horizon values. Various deep learning architectures were tested, based on CNN layers with a supervised learning approach. CNNs were chosen instead of LSTMs due to time and computational constrains as LSTMs have significantly higher training times. History or steps_back and horizon were both defined equal to 24 which means that predictions concern the next day and the historical data used for the predictions concern only the previous day. The way that the data are transformed from a simple dataframe to the proper form for this methodology is described in Section 3.6.4.

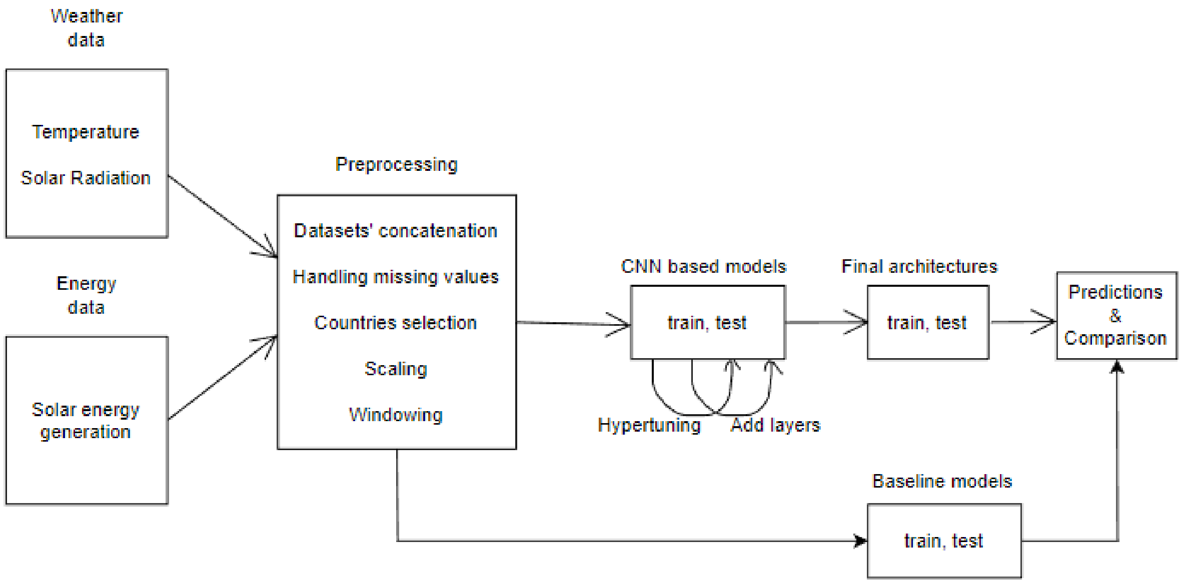

Figure 49: The methodology of this thesis

In Figure 49 the methodology of this thesis is presented in discrete steps. Two different datasets were used and several preprocessing techniques were applied before ending up with the final dataset, as described analytically in the previous sections. Subsequently, the final dataset is given to simple CNN based models for training and testing. Hypertuning techniques was applied to each of the different neural network architectures so as to tune their parameters and to add a proper number of new layers of different types. At the same time, multiple tests took place for each of them so as to identify patterns on the errors and the losses and finally to end up to the best architectures for this task. Each iteration involved a training phase on the training set and an evaluation of the respective model's performance against a validation set. The best architectures are finally used to make predictions for new unknown data and are compared with other baseline models. The metric used to evaluate the models' performance through the process of building the neural networks architectures was the Root Mean Squared Error (RMSE) which is given by the following equation:

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \check{y}_j)}$$

Last but not least, is has to be mentioned that the results presented in this work have been produced using the Aristotle University of Thessaloniki (AUTh) High Performance Computing Infrastructure and Resources [45]. More specifically, after the preprocessing, for the code's execution during the modeling and testing phases, Jupyter notebook interactive application was used via the Aristotle Cluster. It is estimated that for the execution of these tasks, more than 6.500 CPU hours were utilized in the cluster, to train over 250 models.

4.3.1 Simple CNNs Architectures

After the dataset preparation, as described earlier, a simple CNN neural network model was created and initialized. Training was conducted on the selected dataset with a maximum of

200 epochs, and with the addition of an early callback that monitored the validation error, stopped the training when there was no further improvement. This model consisted of one convolutional input layer with ReLU as activation function, followed by a MaxPooling layer with pool size equal to 2, a Flatten layer and finally an output Dense layer with 24 units. It is worth mentioning that the number of units in the output layer is always equal with the forecast horizon. For the previous model, 5 different variations were created based on kernel size parameter (2, 3, 4, 5, 6). For each kernel size, experiments took place for filters number=32, 64, 128 and pool size=2, 3. A learning rate scheduler was used for all models with initial learning rate=1e-2, decay steps=40000 and decay rate=0.9. Also, batch normalization and dropout between the layers were tried. The latter shown in almost all cases that can reduce the RMSE value even in half. In Table 6 an indicative example of the aforementioned process and its results are given.

Table 6: Experiments' results on 1st variation of simple CNN architecture

		RMSE per Country							
Model number	**Characteristics**	**BE**	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**
1	kernel_size = 2 | filters = 64 |pool_size = 2	579	153	1459	2996	27	370	135	91
2	Model 1 with BatchNormalization & Dropout between layers	232	58	346	755	7	158	40	36
3	Model 1 with filters=32	579	155	1461	3023	27	369	135	90
4	Model 3 with BatchNormalization&Dropout between layers	239	59	381	831	6	154	43	36
5	Model 1 with filters=128	579	155	1460	3020	27	370	135	90
6	Model 5 with BatchNormalization&Dropout between layers	230	57	338	729	7	153	40	35
7	Model 6 with pool_size=3	231	59	349	752	8	166	41	36

As shown in Table 6 model number six seems to have the best performance considering the RMSE value for each country. For each of the five variations, the above experiments took place and led to the best five combinations of parameters for each kernel size value. It is clearly seen that RMSE values between some models and their variations were very close (e.g., models 6 and 7). For this reason, for each of the five best variations some statistics were applied, so as to ensure that their performance was indeed at these levels. More specifically, for each variation the model was trained 10 times and the mean of their RMSE scores were calculated for each country. In Table 7 the five best variations with their mean RMSE value for each country are presented. The statistical tests proved that the performances of these 5 models did not appear by chance. It seems that regardless the kernel size value, 128 filters, pool size 2 and batch normalization and dropout between layers are the best parameters combination, with kernel size 5 being the variation which achieved the best score overall.

Table 7: Best parameters combination for each kernel size and their mean RMSE

		Mean RMSE per Country							
Model number	**Characteristics**	**BE**	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**
1	kernel_size = 2 | filters = 128 |pool_size = 2 | BatchNorm & Dropout	230	57	338	729	7	153	40	35
2	kernel_size = 3 | filters = 128 |pool_size = 2 | BatchNorm & Dropout	225	56	339	755	7	150	40	34
3	kernel_size = 4 | filters = 128 |pool_size = 2 | BatchNorm & Dropout	229	57	339	737	7	155	40	34
4	kernel_size = 5 | filters = 128 |pool_size = 2 | BatchNorm & Dropout	222	55	322	699	7	149	39	33
5	kernel_size = 6 | filters = 128 |pool_size = 2 | BatchNorm & Dropout	229	56	342	754	6	152	40	35

4.3.2 Fine Tuning and Adding more Dense layers

In the previous subsection, the first general experiments were presented, based on a very simple CNN neural network architecture. The first important insight given, was that batch normalization and drop out should be used, at least between dense layers. Concerning the kernel size and the number of filters, a more extensive investigation was considered advisable to be done. For this reason, the Optuna framework was used [46]. Optuna is an open-source tool used for parameter optimization and automatic hyperparameter search. For the fine-tuning experiments which are presented in this and the next section, only three out of 8 of dataset's countries were used for the training processes. This happened due to time and computational constrains as the neural networks from now on start to getting deeper, making the training gets slower and slower. At the same time, for each Optuna experiment at least 50 trainings (trials) took place, which also demand extremely many hours and capable computational power.

For this thesis and at this step, Optuna was used to find the best combination of filters and kernel size of the input Convolutional layer and at the same time to create as many dense layers as needed after the first three fixed layers (Convolutional, MaxPooling, Flatten). Also, except for the number of additional dense layers, the number of their units was also included as hyperparameter. Optuna receives a list of hyperparameters and the search space for each parameter, executes as many trials as it is asked and returns a study object with the results. Each trial is a neural network training and validation process, with different combinations of the given hyperparameters and different architectures as different number of layers are added on each trial. The best trial is the one which returns the lowest RMSE countries' average, as defined. Table 8 shows the hyperparameters and their respective search spaces that were given to Optuna.

Table 8: Hyperparameters and their Search Spaces from first Optuna tests

Hyperparameter	Search Space
Filters	[8, 128, 32]
Kernel size	[2, 6, 1]
Dense layers	[3, 20, 3]
Units per layer	[50, 300, 100]

Table 9 presents the characteristics of the three best architectures that arose after the first Optuna tests' execution. Considering the findings of the previous tests, it was decided to try to add batch normalization and dropout between the dense layer of the three best architectures. The scores of each neural network before and after batch normalization and dropout addition are given as well. These scores represent the countries' average RMSE for each neural network architecture. The version with batch normalization and dropout between dense layers is the one chosen for the next steps for all of the three architectures, as it is clearly seen that it leads to lower RMSE.

Table 9: Three best architectures arose from first Optuna tests

Model number	Architecture	Mean RMSE before BatchNorm & Dropout	Mean RMSE after BatchNorm & Dropout
1	Filters: 104 Kernel size: 3 Dense layers: 3 Units per layer: 150	866	194
2	Filters: 72 Kernel size: 3 Dense layers: 6 Units per layer: 250	865	191
3	Filters: 128 Kernel size: 3 Dense layers: 1 Units per layer: 50	728	201

4.3.3 Adding more Convolutional Layers

After ending up to the three architectures referred in Table 9, the Optuna tool was used again to ascertain how many Convolutional layers is worth adding in each architecture that arose from the first tests. Unlike before, this time three different Optuna studies were designed, one for each architecture. The parameters and the layers that arose from the previous steps stay as they are and Optuna checks how many Convolutional layers will be added in a row at the beginning of each neural network and the number of the filters and the kernel size for these layers. Optuna will return the best combination of hyperparameters and layers concerning the mean RMSE of all countries as before. Table 10 gives the hyperparameters and their search spaces as they given to the three different Optuna tests.

Table 10: Hyperparameters and their Search Spaces from second Optuna tests

Hyperparameter	Search Space
Filters	[8, 128, 32]
Kernel size	[2, 6, 2]
Convolutional layers	[3, 10, 3]

Each Optuna study evaluated 45 model variations. The best parameter combinations with the ideal number of Convolutional layers are presented in Table 11, accompanied with their respective mean RMSE score. In all cases, 9 seems to be the ideal number of Convolutional layers at the beginning of the network, with 104 filters in all cases as well. Comparing the last columns of Tables 9 and 11 it is clearly seen that the addition of the Convolutional layers at the beginning of the networks led to an important reduction of RMSE.

Table 11: Three best architectures arose from second Optuna tests

Model number	Architecture	Mean RMSE
1	Filters: 104 Kernel size: 2 Convolutional layers: 9	119
2	Filters: 104 Kernel size: 4 Convolutional layers: 9	140
3	Filters: 104 Kernel size: 6 Convolutional layers: 9	118

4.3.4 Residual Layers, Learning Rate and Final Architectures

In the previous three subsections, the way of building three effective deep neural networks starting from a very simple CNN architecture was presented analytically. After multiple tests and application of fine-tuning techniques in a repetitive manner, we ended up with three different CNN-based neural network architectures, with multiple types of layers and different hyperparameter values. In the last part of the modeling phase, an attempt was made to add Residual layers to the networks, improve their learning rates and try different activation functions for some layers.

More specifically, after different experiments, two Residual layers were added to both models 2 and 3 as shown in Figure 51 and Figure 52. On the other hand, Residuals didn't seem to help model 1 to improve its performance. Additionally, concerning the activation function of dense layers, all of the aforementioned experiments have been done using the Linear function, which is the default activation function for dense layers. However, in the final phase, several experiments took place using the Relu function and this seemed to enhance the performance of model 2. Thus, the activation function of dense layers changed only in the case of the second model.

Figure 50, Figure 51 and Figure 52 show the final three architectures diagrammatically. Inside the parentheses, the number of filters per convolutional layer and the units of each dense layer are given. It is reminded that the output of each architecture resulting from a dense layer with 24 units, which is equal with the horizon size. Table 12 presents the performances of the models. The Normalized RMSE (NRMSE) for each country is given, alongside with countries' overall NRMSE for each model and also the Normalized Mean Absolute Error (NMAE) for each model and each country which was calculated and used for more detailed comparison with the baseline models which are presented in the next section. MAE is given by the following equation:

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|y_j - \check{y}_j\right|$$

The errors are given in a normalized form for easier comparison, which was obtained by dividing the true error value for each country by the mean real value of solar generation (given in Table 4) of the same country.

Table 12: RMSE and MAE scores of final three neural networks

NRMSE per Country									
	BE	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**	**Overall**
Pretrained 1	0.217	0.204	0.139	0.141	0.375	0.205	0.175	0.186	0.205
Pretrained 2	0.247	0.236	0.146	0.134	0.5	0.242	0.164	0.203	0.234
Pretrained 3	0.266	0.258	0.184	0.190	0.5	0.242	0.216	0.237	0.241
NMAE per Country									
Pretrained 1	0.137	0.129	0.094	0.097	0.25	0.130	0.113	0.118	0.133
Pretrained 2	0.129	0.129	0.008	0.007	0.25	0.121	0.092	0.110	0.105
Pretrained 3	0.175	0.172	0.130	0.136	0.312	0.172	0.154	0.135	0.160

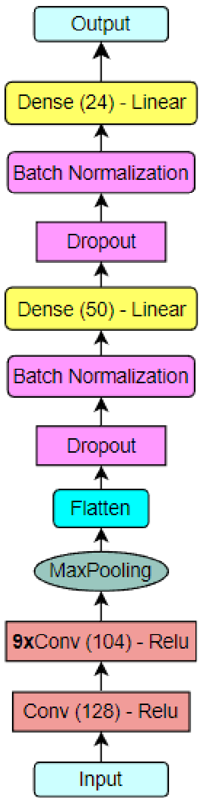

Figure 50: Model 1 architecture

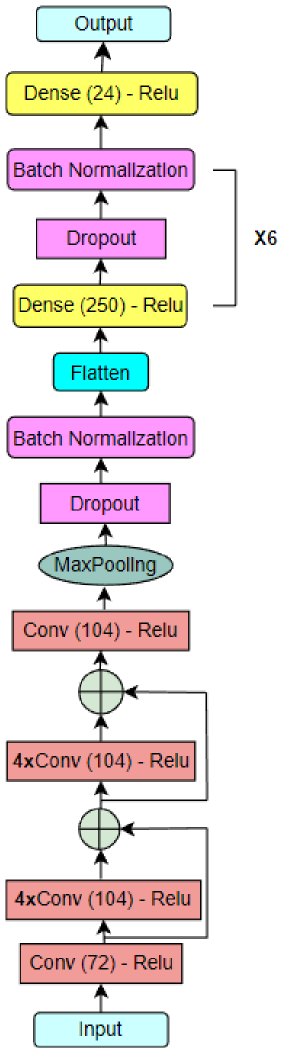

Figure 51: Model 2 architecture

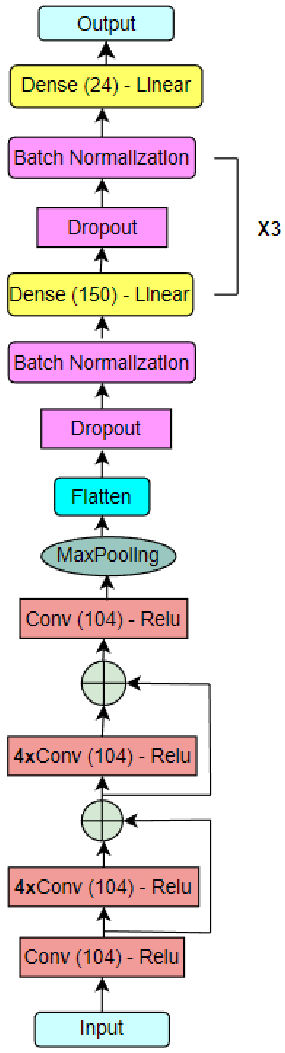

Figure 52: Model 3 architecture

Concerning the learning rate, after ended up to the three final architectures, the focus turned to finding a way to reduce training time, which was very high at that moment. For this task, a different learning rate scheduler was used and several trials were performed for each of the three architectures, with different combinations of scheduler's parameters each time. Through learning rate experimentations, the training time was reduced by 30 epochs maximum but at the same time models' performances also decreased significantly. For this reason, the initial learning rate scheduler was decided to be kept in order to preserve models' high performances.

In general, all three architectures are different combinations of the same layer types, with different hyperparameter values. It seems that model 1 is the best among the three models regarding to RMSE scores, while model 2 seems to have the best performances concerning the MAE scores. Model 3 in both cases is the least efficient architecture. Model 2 is the deepest architecture (35 layers), while model 1 is the shallowest one (18 layers) among the three models, 8 layers less than model 3. All three models are considered as deep architectures, which is the main reason that training time were so high, as a model's performance increases when the depth of the network is higher. In all cases, the addition of the dropout layers improved the training time to some extent. Model 1, compared to the other two models, has just one dense layer before the output and that dense layer has less units (50) than dense layers of the other two models have (250 & 150). Residual layers seem to be useful when the neural networks become very deep, as their addition improved models 2 and 3 performance but not model's 1. Between models 2 and 3, the one with the most dense layers and with the most units within them seems to be the most efficient (model 2). In Figure 53 a 24-hour prediction of each model is presented for Lithuania's solar generation for a randomly selected day within the test set.

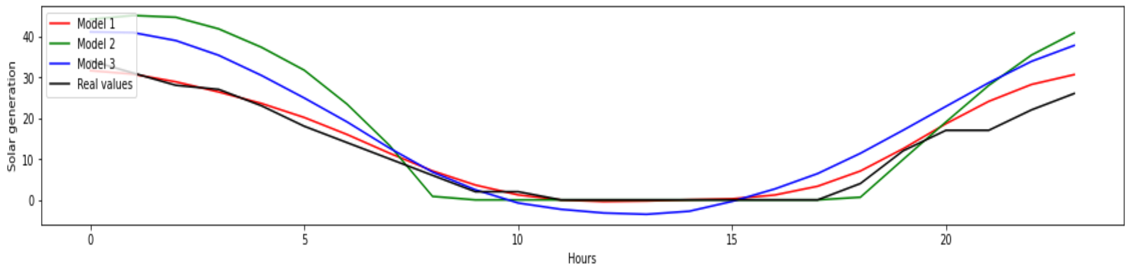

Figure 53: 24-hour predictions for Lithuania from the three neural networks

4.3.5 Baselines Models

The goal of this thesis is to develop neural network architectures trained in massive amount of data from different countries, which can outperform other approaches trained in less data from one country. For the development and the execution of the baseline models, Darts library was used [47]. The Darts library is a user-friendly, open-source library -written in Python for time series forecasting and anomaly detection. It features a range of time series forecasting models, including traditional ones such as Exponential Smoothing, as well as deep neural networks. The goal of Darts is to simplify and improve the end-to-end process of time series machine learning. Each of the models used in this stage, was trained and tested in each country separately and evaluated based on RMSE and MAE metrics.

A backtesting method, which is a method that tries to imitate the windowing method as it makes multiple trainings and predictions in a moving training set based on forecasting_horizon and steps_back parameters was used in all baseline models. Forecasting_horizon and steps_back were both set equal to 24, the same as in the training and testing processes of the neural networks. The backtesting method returns the mean score of the metric which was given. Finally, it should be mentioned that for the baseline models, the only feature that was used for the training process was the solar generation, which is also the one that we want to predict. Solar radiation and temperature, were not used as extra features in the case of most of the baselines, as these models can receive only one input.

The models selected as baselines are the following:

1. Exponential Smoothing
2. Theta Method
3. NaiveMean

4. NaiveDrift
5. Simple CNN

Except for the statistical models of Darts library, a simple CNN neural network was used as a baseline too. This model consists of the same types of layers as the pretrained ones and receives all of the dataset's features. Concerning the statistical baselines, Exponential Smoothing was introduced in Section 2.4.2 and it is considered as one of the most widely used statistical approaches for time series forecasting problems. The forecasts produced from Exponential Smoothing methods are weighted averages of past observations, with the weights getting reduced in an exponentially manner as the observations get older. A fine-tuning procedure was followed for each country with the use of Optuna library again, so as to find the best combination of Exponential Smoothing hyperparameters through which the RMSE score get minimized.

The same procedure was followed for the Theta method as well, in order to find out the best value of theta parameter for each country. The Theta model, introduced by Assimakopoulos and Nikolopoulos in 2000 [48], is a simple forecasting method which consists of fitting two theta lines, forecasting them using the Simple Exponential Smoothing method, and finally combining the forecasts from both lines to generate the final prediction. Finally, NaiveDrift and NaiveMean are two benchmark approaches of Darts library which were used to frame the previous two models. NaiveMean is a non-parametric model which simply predicts the mean value of the training timeseries, while NaiveDrift works by fitting a line between the initial and final point of the training set and projecting it into the future. Table 13 presents the performances of the baseline models concerning NRMSE and NMAE errors for each country.

Table 13: Baselines' performances

NRMSE per Country									
	BE	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**	**Overall**
Exponential Smoothing	0.568	0.559	0.305	0.280	0.625	0.696	0.608	0.389	0.503
Theta Method	0.939	5.129	0.802	1.585	71.25	1.598	4.845	8.432	11.87
Simple CNN	0.623	0.602	0.343	0.383	0.625	0.685	0.453	0.559	0.534
NaiveMean	1.449	1.483	1.343	1.392	1.375	2.485	1.329	0.576	1.366
NaiveDrift	2.502	2.440	2.549	2.011	1.875	2.485	2.484	2.559	2.35
NMAE per Country									
Exponential Smoothing	0.321	0.322	0.173	0.213	0.5	0.397	0.525	0.203	0.331
Theta Method	0.618	3.258	0.514	1.362	44.75	1.052	3.061	5.338	7.494
Simple CNN	0.359	0.397	0.236	0.266	0.375	0.369	0.319	0.355	0.334
NaiveMean	1.217	1.215	1.139	1.202	1.25	2.209	1.546	0.288	1.209
NaiveDrift	2.263	2.193	2.274	1.782	1.75	2.209	2.185	2.322	2.122

Based on the above results and according to both metrics, it is obvious that Exponential Smoothing and simple CNN are by far the best models compared to the other three approaches. The results of both models are very close to each other with Exponential Smoothing outperforming simple CNN in overall score. Comparing the bests of the baseline models with the three neural networks that were constructed, it is clearly seen that the deep neural networks outperform Exponential Smoothing and simple CNN according to both metrics. Pretrained models are better in overall score and also individually in predictions for each of these countries. It is worth mentioning that this comparison concerns predictions in the countries which belong to the dataset on which the pretrained models were trained and not in unknown countries.

As shown in Table 14, model 1 which is the best pretrained architecture based on RMSE metric achieves an overall score 2.5 times smaller compared to Exponential Smoothing overall score and almost 3 times smaller than simple CNN overall score. Observing the predictions per country, there are many cases such as Portugal, Netherlands, Denmark and Belgium, where the pretrained neural network achieves a performance almost 3 times greater than the two baselines. The same applies if we compare the two baselines with the pretrained model 2, which is the best architecture regarding the MAE metric. Overall model's 2 NMAE score is almost 3 times better than the score that the Exponential Smoothing and simple CNN achieve, and in some cases such as Portugal the error is reduced by 5 times when using model 2 instead of the baselines' best models.

Additionally, even the least efficient pretrained neural network outperforms Exponential Smoothing and simple CNN in this category, according to both metrics. Model 3 has overall scores 0.241 and 0.160 for NRMSE and NMAE respectively, when Exponential Smoothing achieves 0.503 and 0.331, and simple CNN 0.534 and 0.334. For all of the countries, baselines have errors at least 2.5 greater than model 3 errors. It is worth mentioning that in all cases, NMAE seems to have lower prices than NRMSE with the level of decline varying between countries. In conclusion, pretrained models seem to be more efficient and outperform by far the baseline ones, when they predict for countries that belong to the dataset they were trained on.

Table 14: Comparison between best pretrained models, Exponential Smoothing and simple CNN in dataset's countries

NRMSE per Country									
	BE	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**	**Overall**
Pretrained 1	0.217	0.204	0.139	0.141	0.375	0.205	0.175	0.186	0.205
Exponential Smoothing	0.568	0.559	0.305	0.280	0.625	0.696	0.608	0.389	0.503
Simple CNN	0.623	0.602	0.343	0.383	0.625	0.685	0.453	0.559	0.534
NMAE per Country									
Pretrained 2	0.129	0.129	0.008	0.007	0.25	0.121	0.092	0.110	0.105
Exponential Smoothing	0.321	0.322	0.173	0.213	0.5	0.397	0.525	0.203	0.331
Simple CNN	0.359	0.397	0.236	0.266	0.375	0.369	0.319	0.355	0.334

The second part of experimentations, concerns the performance of the pretrained models if they were not pretrained. In other words, the same three neural network architectures without any prior knowledge were trained in each of the 8 countries separately and made individual predictions. Table 15 presents the detailed results between the pretrained models and the same architectures trained in each country. It is clearly seen that pretrained models outperform the individually trained models in overall score and in each country separately. Every pretrained model is more efficient than its version trained individually per country and also better than the other two pretrained models' less trained versions. The error reduction in the overall scores between the pretrained models and their individually trained versions is at least 30% in all cases. In general, when it comes to predictions for countries belonging to the dataset that the pretrained models have been trained, it seems that pretrained neural networks predict more accurately compared to the predictions they would make if they were trained in each country individually. That means that the more data for training, the more efficient the neural network will be for this task.

Table 15: Comparison between pretrained models and the same architectures in dataset's countries

NRMSE per Country									
	BE	**DK**	**FR**	**IT**	**LT**	**NL**	**PT**	**SK**	**Overall**
Pretrained 1	0.217	0.204	0.139	0.141	0.375	0.205	0.175	0.186	0.205
1 trained to one country	0.324	0.397	0.291	0.351	0.5	0.452	0.299	0.288	0.362
Pretrained 2	0.247	0.236	0.146	0.134	0.5	0.242	0.164	0.203	0.234
2 trained to one country	0.654	0.677	0.674	0.837	0.75	0.654	0.484	0.610	0.542
Pretrained 3	0.266	0.258	0.184	0.190	0.5	0.242	0.216	0.237	0.241
3 trained to one country	0.331	0.333	0.240	0.245	0.437	0.378	0.216	0.355	0.316
NMAE per Country									
Pretrained 1	0.137	0.129	0.094	0.097	0.25	0.130	0.113	0.118	0.133
1 trained to one country	0.197	0.225	0.178	0.216	0.25	0.242	0.195	0.169	0.209
Pretrained 2	0.129	0.129	0.008	0.007	0.25	0.121	0.092	0.110	0.105
2 trained to one country	0.346	0.344	0.040	0.474	0.437	0.308	0.257	0.338	0.323
Pretrained 3	0.175	0.172	0.130	0.136	0.312	0.172	0.154	0.135	0.160
3 trained to one country	0.216	0.225	0.169	0.171	0.250	0.238	0.164	0.237	0.221

The last part of experimentations, concerns the efficiency of the pretrained models to make predictions for totally unknown countries. For this task, 4 extra countries were selected (Switzerland - CH, Austria - AT, Bulgaria - BG and Czech Republic - CZ) from the initial dataset which was consisted of 20 countries, as described in Section 4.2.3. As shown in Table 16, pretrained models achieved better performances in Bulgaria and Czech Republic, concerning both metrics. On the other hand, in the case of Switzerland, baselines outperform all of the three pretrained models, while in the case of Austria the results are more balanced. More specifically, both baselines can predict more accurately for Austria regarding NRMSE metric, but when it comes to NMAE pretrained models 1 and 2 outperforms simple CNN, which performance is equal to pretrained model 3. Focusing on overall error of each model, pretrained models outperform simple CNN regarding both metrics and Exponential

Smoothing according to MAE. One the other hand, Exponential Smoothing is more efficient in overall based on RMSE metric. In general, the results seem to be balanced and totally related with each country's data, which does not allow us to draw certain conclusions. It is obvious that more experimentations and statistical tests are required, in order to be able to conclude that pretrained models are more efficient than baselines, or the opposite.

Finally, comparing the three pretrained models to each other -for their efficiency to predict for unknown countries, the ranking remains the same as the previous tasks. Model 1 seems to be the most efficient one in overall score and in each country individually -concerning RMSE metric, while model 2 is the most accurate one regarding MAE metric. Model 3 is the least efficient one according to both metrics.

Table 16 shows the errors from both pretrained and baseline models, concerning their predictions in these 4 countries. Also, the mean real value of solar generation for each country is given too. Exponential Smoothing and simple CNN were trained and made predictions for each of these countries separately, when pretrained models made predictions by using their prior knowledge and without any kind of training on the data of these countries.

As shown in Table 16, pretrained models achieved better performances in Bulgaria and Czech Republic, concerning both metrics. On the other hand, in the case of Switzerland, baselines outperform all of the three pretrained models, while in the case of Austria the results are more balanced. More specifically, both baselines can predict more accurately for Austria regarding NRMSE metric, but when it comes to NMAE pretrained models 1 and 2 outperforms simple CNN, which performance is equal to pretrained model 3. Focusing on overall error of each model, pretrained models outperform simple CNN regarding both metrics and Exponential Smoothing according to MAE. One the other hand, Exponential Smoothing is more efficient in overall based on RMSE metric. In general, the results seem to be balanced and totally related with each country's data, which does not allow us to draw certain conclusions. It is obvious that more experimentations and statistical tests are required, in order to be able to conclude that pretrained models are more efficient than baselines, or the opposite.

Finally, comparing the three pretrained models to each other -for their efficiency to predict for unknown countries, the ranking remains the same as the previous tasks. Model 1 seems to be the most efficient one in overall score and in each country individually -concerning RMSE metric, while model 2 is the most accurate one regarding MAE metric. Model 3 is the least efficient one according to both metrics.

Table 16: Predictions for unknown countries

NRMSE per Country					
	CH	**AT**	**BG**	**CZ**	**Overall**
Pretrained 1	1.048	0.436	0.556	0.775	0.703
Pretrained 2	1.121	0.460	0.577	0.795	0.738
Pretrained 3	1.146	0.460	0.583	0.826	0.753
Exponential Smoothing	0.634	0.333	0.684	0.974	0.656
Simple CNN	0.829	1.310	0.678	0.867	0.921
Mean real value	41	126	149	196	
NMAE per Country					
Pretrained 1	0.537	0.230	0.289	0.403	0.364
Pretrained 2	0.488	0.238	0.282	0.398	0.351
Pretrained 3	0.610	0.262	0.322	0.454	0.412
Exponential Smoothing	0.366	0.198	0.590	0.832	0.496
Simple CNN	0.439	0.262	0.550	0.480	0.432
Mean real value	41	126	149	196	

Chapter 5

Conclusions and Future Work

In this dissertation, an introduction on the field of Machine Learning was given with particular emphasis on its relation with the energy sector and specifically with renewable energy sources. The importance of finding solutions to optimize the exploitation of renewable energy units is analyzed, with one of them being the main subject of the thesis: energy production forecasting. In machine learning, energy forecasting is modeled as a time series prediction problem, which can be extremely complicated through the multiple features which may be involved and the quality and the nature of the data that are available for the models.

The data used in this thesis concern hourly measurements from solar generation units for several European countries and for a 5-year span. Also, temperature and solar radiation measurements for these countries were used too. The preprocessing phase proved more challenging than expected, due to existence of missing values in most of the countries and in different periods through the dataset. Certain techniques were used for the data cleaning so as to keep the data in the same order, which is vital when it comes to timeseries data. Preprocessing of the data also included concatenation of the two datasets, adding calendar features and the application of a windowing method.

The experiments that took place focused on the construction of efficient Convolutional neural networks architectures which make 24-hour ahead predictions on the solar generation data. Several trials took place for different combinations of layers and hyperparameter values with the total time of experiments reaching almost 6.500 CPU hours. The experiments flow was based on the reduction of RMSE metric and ended up with three deep neural network architectures mainly composed of Convolutional and Dense layers. Among the three models, the one with the least dense layers was the most efficient concerning RMSE metric, while the deepest network was the best one regarding MAE metric. It is worth mentioning that the training time was very high for all of the three deep neural networks. The addition of Dropout and BatchNormalization between the layers seemed to be helpful in all cases, while residual layers were beneficial only for the two of the three models.

In parallel, four statistical models and one simple Convolutional Neural Network were trained individually to each country and used for comparison with the pretrained models. Exponential Smoothing and simple CNN were the best models within the baselines and they were the ones which was selected for further experimentation and comparison. All of the three neural networks outperform by far the baselines and achieved more accurate predictions for the countries which belong to the dataset that were trained on. The error reduction between the pretrained neural networks and the baselines even reached 300% in some cases. Also, it was observed that pretrained neural networks can outperform neural networks with exactly the same architectures, when the first are trained in massive volumes of data -among them the data of the country that they are trying to predict, and the latter are trained on just this country's data.

In the last part of experimentations, pretrained models were tested in totally unknown data. Predictions concern 4 new countries on which baseline models were trained individually to each of them before make predictions, while pretrained models made predictions for them using their prior knowledge and without any kind of training on new countries' data. Overall scores of pretrained models were better than both baselines' overall scores according to MAE metric, but worse than Exponential Smoothing overall score regarding to RMSE. The results were balanced and seems to be related with the data of each country. That means that more experimentations and statistical tests should be done, so as to conclude that pretrained models are more efficient than baselines, or the opposite.

Regarding the future work, more countries could be added to the training process of the neural networks and extra features as well, so as for the model to produce more informed results. Future weather predictions could be a beneficial extra feature for neural networks if added. Additionally, the efficiency of the models could be tested in different horizons except for 24 and with different sizes of windows in the windowing process. Fine-tuning the hyperparameters of each model for each country is also an interesting approach which is worth testing in the effort to further improve the accuracy of their predictions.

Furthermore, in order to enhance the performance of pretrained models in unknown countries, freezing techniques could be used. Pretrained models could freeze all of their layers -except for the final layer which will be used for some training with the unknown data, while the rest of the layers will keep their weights frozen. This will add some extra information for the pretrained models which could be proved vital, without requiring them to be trained from scratch. Finally, as mentioned in Section 4.3.4, the training time for all of the three neural networks was very high due to great depth that all three architectures have. It would be beneficial to find a learning rate scheduler which can reduce the training time and at the same time maintain the high performance of the models.

References

[1] W. Huaizhi , . L. Zhenxing, Z. Xian and Z. Bin, "A review of deep learning for renewable energy forecasting," *Elsevier,* 15 10 2019.

[2] S. Russel and P. Norvig, Artificial Intelligence : a Modern Approach, 2010.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning, vol. 20,* pp. 273-297, 1995.

[4] C. Kingsford and S. Salzberg, "C. Kingsford and S. L. Salzberg, "What are decision trees?,"," *Nature Biotechnology,* pp. 1011-1013, 9 2008.

[12 S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory, vol 28,* pp. 129-137, 3 1982.

[6] L. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications, USA, 1994.

[7] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv,* 2017 7 2017.

[8] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *arXiv,* 2015.

[9] J. L. Elman, "Finding Structure in Time," *Cognitive Science 14,* pp. 179-211, 3 1990.

[10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation 9,* 15 11 1997.

[11] K. Cho, B. van Merrienboer, D. Bahdanau and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," *arXiv,* 2014.

[12] S. Richard and N. Shirish Keskar, "Improving Generalization Performance be Switching from ADAM to SGD," *arXiv,* 12 2017.

[13] M. Qi and G. P. Zhang, "Trend Time–Series Modeling and Forecasting With Neural Networks," *IEEE Trans. Neural Networks,* vol. 5, pp. 808-816.

[14] D. Ladiray and B. Quenneville, "Outline of the X-11 Method. In: Seasonal Adjustment with the X-11 Method," in *Lecture Notes in Statistics*, New York, 2001.

[15] A. Raudys, V. Lenčiauskas and E. Malčius, "Moving averages for financial data smoothing," *Commun. Comput. Inf. Sci.,* pp. 34-45, 2013.

[16] Statistical Review of World Energy BP p.l.c., "bp," 2021. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf.

[17] E. Commision, "Energy Data Tables," 2020.

[18] H. Ritchie, M. Roser and P. Rosado, "Energy," *Our World in Data,* 2022.

[19] E. Comission, "Where does our energy come from?," 2020.

[20] E. Comission, "EU energy mix and import dependency," 2022.

[21] E. F. Moran, M. Claudia Lopez, N. Moore and D. W. Hyndman, "Sustainable hydropower in the 21st century," *PNAS,* 5 11 2018.

[22] "IPCC," [Online]. Available: https://www.ipcc.ch/.

[23] "Kyoto Protocol Reference Manual on Accounting of Emissions and Assigned Amounts," UNFCCC, 2008.

[24] E. Comission, "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS," 2020.

[25] Hellenic Association for Energy Economics, "haee," 2022. [Online]. Available: https://www.haee.gr/media/4549/gemr-2022_final.pdf.

[26] M. Mastorakis, "ENERGY MATTERS TO CLIMATE CHANGE," 2020.

[27] energypedia, "Greece Energy Situation," 2021.

[28] HELLENIC REPUBLIC - Ministry of the Environment, *National Energy and Climate Plan,* 2019.

[29] Y. Shengxi , A. S. Kocaman and M. Vijay, "Benefits of forecasting and energy storage in isolated grids with large wind penetration – The case of Sao Vicente," *Elsevier,* 5 2017.

[30] A. Alzahrani, P. Shamsi , C. Dagli and M. Ferdowsi, "Solar Irradiance Forecasting Using Deep Neural Networks," *Elsevier,* 2017.

[31] H. Pedro and C. Coimbra, "Short-term irradiance forecastability for various solar micro-climates," *Elsevier,* pp. 587-602, 12 2015.

[32] Z. Jinsong, V. Rodrigo, N. Shohei and L. Jean-Francois, "Deep photovoltaic nowcasting," *Elsevier,* pp. 267-276, 12 2018.

[33] C. Timplalexis, N. Bezas, A. Bintoudi and L. Zyglakis, "A hybrid physical/statistical day-ahead direct PV forecasting engine," *ResearchGate,* 11 2020.

[34] C. Jie, Z. Guo-Qiang, Z. Wuneng, D. Wei and L. Kang-Di, "Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization," *Elsevier,* pp. 681-695, 6 2018.

[35] Z. M. Hamza, K. N. Mujeed, M. Majad, M. A. Feroz, M. S. Raza and F. Sanfilippo, "Adaptive ML-based technique for renewable energy system power forecasting in hybrid PV-Wind farms power conversion systems," *Elsevier,* 4 2022.

[36] G. R. da Silva, H. M. Dal Molin Ribeiro, R. S. Monero, V. Cocco Mariani and L. dos Santos Coelho, "A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting," *Elsevier,* 2 2021.

[37] F. Shahid, D. Wood, H. Nisar, A. Zameer and S. Eger, "Short-term multi-hour ahead country-wide wind power prediction for Germany using gated recurrent unit deep learning," *Elsevier,* 10 2022.

[38] D. Salinas, V. Flunkert, J. Gasthaus and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Elsevier,* 9 2020.

[39] B. Lim, S. O. Arik, N. Loeff and T. Pfister, "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting," *arXiv,* 19 12 2019.

[40] B. Oreshkin, D. Carpov, N. Chapados and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," *ICLR 2020,* 24 5 2019.

[41] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir and R. Rajagopal, "NeuralProphet: Explainable Forecasting at Scale".

[42] "Open Power System Data," [Online]. Available: https://open-power-system-data.org/.

[43] "Open Power System Data - Energy," 2020. [Online]. Available: https://data.open-power-system-data.org/time_series/2020-10-06.

[44] "Open Power System Data - Weather," 2020. [Online]. Available: https://doi.org/10.25832/weather_data/2020-09-16.

[45] "Aristotelis Docs," [Online]. Available: https://hpc.it.auth.gr/.

[46] A. Takuya, S. Shotaro, Y. Toshihiko, O. Takeru and K. Masanori, "Optuna: A Next-generation Hyperparameter Optimization Framework," *KDD Conference,* 16 7 2019.

[47] J. Herzen, F. Lässig, S. Giuliano Piazzetta, T. Neuer, L. Tafti, G. Raille, T. Van Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. G. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik and G. Grosch, "Darts: User-Friendly Modern Machine Learning for Time Series," *Journal of Machine Learning Research 23,* 19 5 2022.

[48] V. Assimakopoulos and K. Nikolopoulos, "The theta model: a decomposition approach to forecasting," *Elsevier,* 12 2000.