**Σκοπός και Πολιτική Σημασία**

* Ο τουρισμός συνεισφέρει σημαντικά στην ευρωπαϊκή οικονομία (περίπου 10% της απασχόλησης, ~10% του Παγκόσμιου ΑΕΠ και των θέσεων εργασίας). Οι κλιματικές αλλαγές – υψηλότερες θερμοκρασίες, μεταβολές βροχόπτωσης, ακραία φαινόμενα – αναδιαμορφώνουν τις προτιμήσεις ταξιδιού (π.χ. μετεγκατάσταση ζήτησης Βορρά προς Νότο) και απειλούν ευάλωτες τουριστικές οικονομίες.
* Υπάρχει έλλειμμα υφιστάμενων εργαλείων σύνδεσης των κλιματικών δεικτών με τα τουριστικά αποτελέσματα και μεταφοράς γνώσης σε πολιτικές και επενδυτές. Το CRISI στοχεύει στην κάλυψη αυτού του κενού μέσω ενός *AI*-βασισμένου συστήματος αξιολόγησης κινδύνου/ανθεκτικότητας για τον τουρισμό.
* Το εργαλείο θα παράγει *βαθμολογία ανθεκτικότητας* (resilience score) για περιφέρειες ή έργα, βοηθώντας λήψη αποφάσεων σε επίπεδο πολιτικής και επενδύσεων (π.χ. προτεραιοποίηση έργων προσαρμογής στην τουριστική υποδομή). Εναρμονίζεται με τις ευρωπαϊκές στρατηγικές κλιματικής προσαρμογής (π.χ. Ευρωπαϊκό Σχέδιο για την Προσαρμογή) και τους στόχους για βιώσιμο τουρισμό, απαντώντας στην ανάγκη μακροπρόθεσμου σχεδιασμού προσαρμογής στον τουρισμό (π.χ. *SDG* για ανάπτυξη χωρίς αποκλεισμούς).

**AI/Machine Learning Στόχοι**

* **Προβλεπτικό Μοντέλο Επιπτώσεων:** Εκπαίδευση *supervised* μοντέλου που θα προβλέπει αλλαγές στη ζήτηση τουρισμού με βάση κλιματικούς παράγοντες. Π.χ. χρήση Random Forest ή panel regression πάνω σε ιστορικά στοιχεία (τιμές θερμοκρασίας, χιονιού, βροχόπτωσης κ.ά.) για να ποσοτικοποιηθεί ο αντίκτυπος στην τουριστική άφιξη/έσοδα.
* **Δεδομένα Πολυ-μορφικά:** Ο συνδυασμός *multimodal* εισόδων: κλιματικά δεδομένα (θερμοκρασία, υγρασία, ακραία φαινόμενα, δείκτες π.χ. Tourism Climate Index) μαζί με οικονομικά/κοινωνικά δεδομένα (αφίξεις τουριστών, τουριστικά έσοδα, ΑΕΠ περιφέρειας, απασχόληση στον τουρισμό, υποδομές). Αυτό επιτρέπει στο μοντέλο να λαμβάνει υπόψη ευρύτερους παράγοντες τρωτότητας/προσαρμογής.
* **Τεχνικές Μάθησης:** Πιθανή χρήση *machine learning* και στατιστικών (π.χ. *Random Forest*, *XGBoost*, πολυμεταβλητές παλινδρομήσεις ή Bayesian μοντέλα). Εξετάζεται επίσης ο ρόλος *unsupervised* μεθόδων (clustering, PCA) για ανακάλυψη μοτίβων τρωτότητας στις περιφέρειες ή μείωση διαστάσεων. Η επιλογή του αλγορίθμου θα ισορροπήσει απόδοση με διαφάνεια.
* **Explainability (Εξηγήσιμα Μοντέλα):** Δεδομένου του policy-φάσματος χρήσης, το μοντέλο πρέπει να είναι εξηγησιμό (XAI). Επιδιώκουμε αλγορίθμους με **εύκολη ερμηνεία** (π.χ. προβολή coefficients σε γραμμικά μοντέλα ή *feature importances* σε δέντρα απόφασης). Εργαλεία όπως SHAP ή LIME μπορούν να βοηθήσουν στην εξήγηση επιρροής των input μεταβλητών (π.χ. πώς μια επιπλέον ημέρα καύσωνα μειώνει τα έσοδα). Το σύστημα θα επιτρέπει την ανάλυση ευαισθησίας (sensitivity analysis) ώστε να κατανοούνται οι κύριοι οδηγοί της βαθμολογίας.
* **Εκπαίδευση και Βελτιστοποίηση:** Χρήση *cross-validation* και διαχωρισμός εκπαιδευτικών/δοκιμαστικών συνόλων (π.χ. χρονοσειράς ή χωρική διαίρεση) για αξιολόγηση γενίκευσης. Υπολογισμός μετρικών σφάλματος και ρύθμιση υπερπαραμέτρων (hyperparameter tuning) εξασφαλίζουν εγκυρότητα και βελτιστοποιούν την απόδοση.

**Data Architecture Overview**

* **Πηγές Δεδομένων:** Χαρτογράφηση υψηλής ανάλυσης από κλιματικά μοντέλα (*EURO-CORDEX* RCMs ή CMIP6) για μελλοντικά σενάρια (π.χ. RCP4.5, RCP8.5), συμπεριλαμβανομένων μεταβλητών όπως θερμοκρασία, βροχόπτωση, δείκτες ξηρασίας, ύψος χιονιού, επίπεδο θάλασσας. Στατιστικά τουρισμού και οικονομικά (αφίξεις, έσοδα, τουριστικό ΑΕΠ) από Eurostat, εθνικούς φορείς και UNWTO. Επιπλέον, κοινωνικο-οικονομικοί δείκτες (π.χ. ποσοστό απασχόλησης στον τουρισμό, υποδομές) για εκτίμηση ευαισθησίας και ικανότητας προσαρμογής.
* **Προεπεξεργασία και Features:** Ομαδοποίηση – καθαρισμός δεδομένων, αντιμετώπιση ελλείψεων, ευθυγράμμιση χρονικών περιόδων. Χωρική συσχέτιση (spatial join): αντιστοίχιση των κλιματικών πλέγματος με διοικητικές περιφέρειες (π.χ. NUTS2) χρησιμοποιώντας GIS βιβλιοθήκες (*GeoPandas*, *rasterio*). Συντεταγμένες θα προβληθούν σε κοινό CRS (π.χ. ETRS89/Europe) ώστε να επιτευχθεί ομοιογένεια.
* **Feature Engineering:** Δημιουργία προσαρμοσμένων δεικτών: π.χ. *Tourism Climate Index*, ημέρες καύσωνα/παγετού, δείκτης αξιοπιστίας χιονιού (ski reliability index). Οι δείκτες αυτοί συνοψίζουν πολύπλοκα φαινόμενα σε εισόδους μοντέλου. Επίσης, κανονικοποίηση (scaling) και διαχείριση κυκλικών δεδομένων (π.χ. εποχιακότητα μήνα) εξασφαλίζουν σταθερότητα. Τα features τελικά παράγονται σε δεδομένα πίνακα (DataFrame) για το ML pipeline.
* **Delphi-βάρυση ως Hybrid Input:** Τα βάρη που προκύπτουν από τη μέθοδο Delphi (εκτίμηση εμπειρογνωμόνων) θα χρησιμοποιηθούν ως *meta-input* στο μοντέλο ή στο στάδιο υπολογισμού της τελικής βαθμολογίας. Για παράδειγμα, ένας παράγοντας που οι εμπειρογνώμονες θεωρούν κρίσιμο (π.χ. υδάτινοι πόροι) μπορεί να ενισχύσει το αντίστοιχο χαρακτηριστικό ή να προσμετρηθεί με μεγαλύτερο βάρος στο composite score. Αυτός ο «υβριδικός» συνδυασμός δεδομένων και expert input αυξάνει την εγκυρότητα του μοντέλου.
* **Αρχιτεκτονική Pipeline:** Το σύστημα θα ακολουθεί τυπική *ETL* ροή: συλλογή → μετασχηματισμός → φόρτωση. Προτείνεται χρήση Python libraries (Pandas/GeoPandas για tabular και spatial δεδομένα, Xarray ή netCDF4 για κλιματικά δεδομένα, Scikit-learn για ML, PyMC3/Stan για πιθανοτικά μοντέλα). Η αποθήκευση μπορεί να γίνει σε PostGIS ή απλά σε αρχεία (GeoJSON, CSV, NetCDF) με version control. Κάθε στάδιο καταγράφεται σχολαστικά για να διασφαλιστεί αναπαραγωγιμότητα και auditability.

**Πλατφόρμα Ανάπτυξης**

* **Python Stack:** Βασιζόμαστε σε Python και το data science οικοσύστημά της (Pandas, NumPy, Scikit-learn, Xarray, PyMC3, TensorFlow/PyTorch κ.ά.). Για γεωχωρική επεξεργασία χρησιμοποιούνται GeoPandas, Fiona/OGR, Shapely. Εάν η επεξεργασία του κλιματικού πλέγματος είναι βαριά, εργαλεία όπως Dask ή Google Earth Engine θα αξιολογηθούν.
* **Ανάπτυξη Web-εργαλείου:** Στη φάση πρωτοτύπου, προτείνεται το **Streamlit** (open-source Python framework) για γρήγορη δημιουργία interactive dashboards (με χάρτες και controls). Επιτρέπει εμφάνιση του χάρτη Ευρώπης με χρωματική κλίμακα των scores και επιλογές σεναρίου/χρονικού ορίζοντα. Εναλλακτικά, μπορούν να χρησιμοποιηθούν Dash ή Bokeh. Το frontend θα στοχεύει σε απλότητα ώστε πολιτικοί χρήστες να κατανοούν τα αποτελέσματα.
* **Ανάγκες Υποδομής:** Η επεξεργασία χωρικών δεδομένων μπορεί να γίνεται τοπικά ή σε cloud (π.χ. AWS/GCP). Για μεγάλο όγκο κλιματικών δεδομένων, ενδεχόμενη χρήση βάσης Time-series ή RasterDB. Σημειώνονται βέλτιστες πρακτικές όπως χρήση Docker για containerization και GitHub για version control, διευκολύνοντας συνεργασία και μελλοντική κλιμάκωση.

**Δυνατότητα Συνεργασίας σε ML**

* **Μοντελοποίηση και Ανάλυση Δεδομένων:** Ερευνητής με εμπειρία σε *machine learning* και βελτιστοποίηση μπορεί να βοηθήσει στο σχεδιασμό και επιλογή αλγορίθμων (π.χ. Random Forest, Gradient Boosting, μοντέλα Seq2Seq ή GAN εάν χρειαστεί). Ο/η συνάδελφος μπορεί να ασχοληθεί με *feature engineering*, επιλογή μεταβλητών, και διερεύνηση επιπλέον (meta)χαρακτηριστικών όπως πολυπλοκότερος υπολογισμός των δεικτών ή ενσωμάτωση νέων πηγών δεδομένων.
* **Υπερπαραμετροποίηση (Hyperparameter Optimization):** Εμπειρία σε εργαλεία όπως GridSearchCV, Bayesian optimization, θα είναι χρήσιμη για fine-tuning των ML μοντέλων, μεγιστοποιώντας ακρίβεια χωρίς υπερπροσαρμογή.
* **Ερμηνεία και Διαφάνεια:** Οποιοσδήποτε με γνώση XAI frameworks (π.χ. SHAP, LIME, partial dependence plots) μπορεί να υλοποιήσει τη διερεύνηση *feature importances* και να ελέγχει αν οι επιρροές των μεταβλητών συμφωνούν με τις κλιματικές/οικονομικές θεωρίες. Αυτό είναι κρίσιμο για εμπιστοσύνη των πολιτικών στο μοντέλο.
* **Αβεβαιότητα και Σενάρια:** Εξειδίκευση σε στατιστική ή πιθανότητες βοηθά στην ποσοτικοποίηση αβεβαιοτήτων (π.χ. μέσω Bayesian μοντέλων ή ensembles). Η εκτίμηση διακύμανσης βάσει διαφορετικών κλιματικών σεναρίων (RCP) και δεδομένων μπορεί να γίνει με Monte Carlo ή bootstrap μεθόδους.
* **Χωρική Ανάλυση:** Γνώση GIS βοηθά στην επεξεργασία χωρικών δεδομένων (spatial joins, reprojection, raster-vector μετατροπές). Η ενσωμάτωση πολυδιάστατων πηγών (ατμόσφαιρα, ωκεανός, ανθρώπινες δραστηριότητες) απαιτεί χειρισμό γεωχωρικών βιβλιοθηκών.
* **Delphi και Domain Knowledge:** Σύνδεση με εμπειρογνώμονες: συνάδελφος με δεξιότητες σε surveys/Delphi μπορεί να συντονίσει τη διανομή ερωτηματολογίων και την επεξεργασία απαντήσεων. Οι βάσεις γνώσης (experts) «κωδικοποιούνται» ώστε να ενσωματώνονται ως βάρη/κλαδικά χαρακτηριστικά.

**Ερωτήματα προς Συζήτηση**

* **Επιλογή Μοντέλου:** Υπάρχει ανάγκη για συγκριτική αξιολόγηση διαφόρων προσεγγίσεων (Random Forest vs. XGBoost vs. πολυμεταβλητή παλινδρόμηση vs. deep learning). Ποιοι παράγοντες (π.χ. ανάγκη για interpretability) θα καθορίσουν την τελική επιλογή;
* **Όγκος και Απόδοση:** Πόσες περιφέρειες (π.χ. NUTS2) και πόσα έτη δεδομένων θα χρησιμοποιηθούν; Η υπολογιστική πολυπλοκότητα των χωρικών μοντέλων και του μεγάλου πλέγματος κλίματος απαιτεί σχεδιασμό (π.χ. επεξεργασία σε φάσεις ή χρήση cloud).
* **Validation:** Ποιες μέθοδοι cross-validation είναι κατάλληλες σε χωρικά-χρονικά δεδομένα (π.χ. k-fold κατά γεωγραφικές υποπεριοχές ή time series split); Θα χρησιμοποιηθεί backtesting με παρελθούσες κλιματικές συνθήκες για αξιολόγηση προβλέψεων;
* **Πιθανά Κρούσματα Μελέτης:** Προτείνουμε διαφορετικά περιβάλλοντα ως πιλοτικά: Μεσογειακό παράκτιο (π.χ. Ελλάδα, Ιταλία) για υψηλή έκθεση σε καύσωνες/άνοδο στάθμης, Αλπικό/χειμερινό (Αυστρία, Ελβετία) για απώλειες χειμερινού τουρισμού, βόρειο (Σκανδιναβία) για πιθανά οφέλη από πιο ήπια κλίματα. Επιπλέον, μεγάλα αστικά κέντρα (π.χ. Παρίσι, Βαρκελώνη) καθώς και μικρότερες νησιωτικές οικονομίες θα έδιναν ποικιλία σε συνθήκες.
* **Περιορισμοί και Υποθέσεις:** Ποιες είναι οι βασικές υποθέσεις (π.χ. γραμμικότητα επιρροής κλίματος στατιστικών τουρισμού) που πρέπει να ελεγχθούν; Πώς χειριζόμαστε αβεβαιότητες όπως αναντιστοιχίες στις αναφορές δεδομένων ή μη μετρήσιμες παραμέτρους (π.χ. αλλαγές πολιτικής);