



VANESSA KENNICHE SANOCKA

Ingénieur Data Scientist



KENSAN.VAN@GMAIL.COM | +33 7 68 14 01 74 | vanessa-kenniche-sanocka.github.io

Paris – Français natif, Polonais natif, Anglais C1 (TOEIC)

Expertise

Spécialisée en économie quantitative et analyse de politiques publiques. Conduite de projets : modélisation économétrique, traitement de données volumineuses et sensibles, restitution d'analyses complexes en recommandations accessibles aux multi-parties prenantes. Programmation et solutions de visualisation innovantes.

Compétences

Modélisation : séries temporelles, panel data, event study, inférence causale, ML, geospatial analytics

Langages & Data : Python, R, SQL, Dplyr, Pandas, Plotly, Scikit-learn, PyTorch, TensorFlow, Airflow, NLP

Outils : Shiny, Dash, Power BI, Excel, Quarto, Git, Docker, Apache Spark, MySQL, MongoDB, QGIS, CASD

Expérience professionnelle – Cour des Comptes, 2023 - Aujourd'hui

- **Analyse géospatiale transfrontalière franco-suisse (Cour des comptes Suisse)** : Harmonisation de données SIG hétérogènes France-Suisse pour identifier les discontinuités réglementaires transfrontalières. Transformation d'analyses géospatiales en visualisation et recommandations de politique publique aux équipes institutionnelles.
- **Évaluation des politiques tarifaires des transports urbains** : Modèles de panel & event study (données nationales GART, CEREMA) et étude par sondage (échantillonnage sur base confidentielle des usagers, redressement par score de propension, calage sur marges, bootstrapping), enrichie de la base FILOSOFI par recherche complémentaire, afin de formuler des recommandations sur les politiques tarifaires, avec restitution interactive via Shiny.
- **Pipeline NLP et RAG** : Intégration full-stack pour l'exploration et l'interrogation sémantique des documents de la Cour des comptes par indexation Elasticsearch, embeddings, LLM, Sentence Transformers et API vers plateforme Next.js, incluant un portage des données de la Cour vers un nouveau système d'archivage interne.
- **Modélisation prédictive sur registres nationaux confidentiels (CASD)** : Algorithmes ML sur données judiciaires pour analyse de survie et scoring de risque. Modélisation sur données fiscales (EEC) pour décision publique.

Expérience académique et personnelle – Cours Master Ingénierie Data Science for Social Sciences, Université Paris Nanterre, 2020 - 2025

- **Détection fraude bancaire** : Modélisation supervisée sur données déséquilibrées, techniques SMOTE/B-SMOTE, validation croisée pour scoring de risque financier.
- **Détection de pricing algorithmique de marché – avec le laboratoire EconomiX** : Scraping haute fréquence vers datalake, classification et estimation des ajustements de prix par modèles d'effets fixe et tests de robustesse.
- **Benchmarking des déterminants d'innovation (Asie du Sud)** : Modélisation économétrique cross-country (logit/Poisson) sur données d'enquête avec tests de robustesse par régularisation Lasso.
- **Network analysis des votes parlementaires** : Scraping opendata gouvernemental, matrices de distance Manhattan et clusters spring layout pour identifier les coalitions transpartisanes.
- **Pipeline Power BI démographique** : Entrepôt en étoile MySQL, requêtes à fenêtrage (moyennes mobiles), tableau de bord territorial pour indicateurs de politique des logements privés vacants.
- **Analyse brevets PATSTAT** : Cartographie des risques concurrentiels et positionnement stratégique sectoriel avec visualisations statistiques et géospatiales pour recommandations de politique industrielle.

Informations additionnelles

- **Ambassadrice pour le plan STEM "Filles et Maths"** sensibilisation de lycéennes aux carrières scientifiques.
- Conception d'outils formation et plateforme collaborative pour les Data Scientists de la Cour des comptes.