

## Projet : 100 Notebooks d'Ingénieur IA (Orienté à la santé)

### Bloc 1 : Fondations Python & Data Science (Notebooks 1-15)

#	Titre du Notebook	Dataset	Ressources Clés
1	NumPy avancé pour le ML	Données synthétiques	<i>Python for Data Analysis</i> (McKinney), <a href="#">NumPy Docs</a>
2	Pandas pour données médicales	<a href="#">UCI Heart Disease</a>	<i>Pandas Cookbook</i> , Kaggle Learn Pandas
3	Visualisation avec Matplotlib/Seaborn	<a href="#">Diabetes Dataset</a>	<i>Python Data Science Handbook</i> (VanderPlas)
4	Statistiques pour le ML médical	<a href="#">MIMIC-III Demo</a>	<i>Practical Statistics for Data Scientists</i>
5	Web Scraping médical (BeautifulSoup)	PubMed articles	<a href="#">Real Python Web Scraping</a>
6	APIs médicales (Selenium)	<a href="#">ClinicalTrials.gov API</a>	Documentation Selenium
7	Nettoyage de données cliniques	<a href="#">CDC NHANES</a>	<i>Data Cleaning</i> (Ilya Zaitsev)
8	Feature Engineering médical	<a href="#">Breast Cancer Wisconsin</a>	Kaggle Feature Engineering Course
9	Gestion déséquilibre classes	<a href="#">Fraud Detection</a>	<i>Imbalanced Learning</i> (He & Ma)
10	Pipeline Scikit-learn complet	<a href="#">Liver Disorders</a>	<a href="#">Scikit-learn Pipeline Guide</a>
11	Validation croisée stratégies	Données synthétiques	<i>Hands-On ML</i> Ch4 (Géron)
12	Optimisation hyperparamètres (Optuna)	<a href="#">Dermatology Dataset</a>	<a href="#">Optuna Documentation</a>
13	SQL pour données médicales	<a href="#">SQLite Medical DB</a>	<i>SQL for Data Scientists</i>
14	Big Data avec Dask	<a href="#">Large CSV Medical</a>	<a href="#">Dask Tutorial</a>
15	Projet : Pipeline ETL complet	Multiple sources	Intégration Notebooks 1-14

**Bloc 2 : Machine Learning Classique (Notebooks 16-30)**

#	Titre du Notebook	Dataset	Ressources Clés
16	Régression linéaire avancée	<a href="#">Boston Housing</a>	<i>ESL</i> Ch3 (Hastie)
17	Régression Ridge/Lasso/ElasticNet	<a href="#">Medical Cost Personal</a>	<a href="#">Glmnet Vignette</a>
18	Régression logistique médical	<a href="#">Heart Disease</a>	<i>An Introduction to Statistical Learning</i> Ch4
19	SVM pour classification médicale	<a href="#">Cervical Cancer</a>	<i>Pattern Recognition &amp; ML</i> Ch7 (Bishop)
20	Decision Trees & Pruning	<a href="#">Thyroid Disease</a>	Scikit-learn Decision Tree Guide
21	Random Forest pour diagnostic	<a href="#">Parkinson's Disease</a>	<i>Random Forests</i> (Breiman 2001)
22	Gradient Boosting (XGBoost)	<a href="#">Diabetes 130-US hospitals</a>	<a href="#">XGBoost Documentation</a>
23	LightGBM pour données médicales	<a href="#">Maternal Health Risk</a>	<a href="#">LightGBM Paper</a>
24	CatBoost avec features catégorielles	<a href="#">Stroke Prediction</a>	<a href="#">CatBoost Tutorial</a>
25	K-Means pour segmentation patients	<a href="#">Customer Segmentation</a>	<i>Pattern Recognition &amp; ML</i> Ch9
26	DBSCAN & clustering hiérarchique	<a href="#">Gene Expression Cancer</a>	Scikit-learn Clustering Guide
27	PCA & t-SNE visualisation	<a href="#">MNIST</a>	<i>Hands-On ML</i> Ch8
28	Autoencoders classiques	MNIST	<a href="#">Keras Autoencoder Blog</a>
29	Gaussian Mixture Models	<a href="#">Iris Dataset</a>	<i>Pattern Recognition &amp; ML</i> Ch9
30	Projet : Système prédiction risque	Combinaison datasets médicaux	Intégration ML classique

**Bloc 3 : Deep Learning Fondations PyTorch (Notebooks 31-45)**

#	Titre du Notebook	Dataset	Ressources Clés
31	PyTorch basics & tensors	Données synthétiques	<a href="#">PyTorch Official Tutorial</a>
32	Autograd & backpropagation	Données synthétiques	<i>Deep Learning</i> Ch6 (Goodfellow)
33	Dataset & DataLoader custom	<a href="#">Medical MNIST</a>	PyTorch Data Loading Tutorial
34	MLP from scratch PyTorch	<a href="#">Pima Indians Diabetes</a>	<i>Neural Networks from Scratch</i>
35	Optimizers comparison (SGD, Adam, etc.)	MNIST	<a href="#">An Overview of Gradient Descent Optimization</a>
36	Learning rate scheduling	CIFAR-10	<a href="#">PyTorch LR Scheduler</a>
37	Batch Normalization & Layer Norm	CIFAR-10	<a href="#">Batch Normalization Paper</a>
38	Dropout & Regularization	<a href="#">Fashion MNIST</a>	<i>Deep Learning</i> Ch7
39	Weight Initialization strategies	Données synthétiques	<a href="#">Xavier/He Initialization Papers</a>
40	Early Stopping & Model Checkpointing	<a href="#">Skin Cancer MNIST</a>	PyTorch Best Practices
41	Transfer Learning basics	<a href="#">ChestX-ray14</a>	<a href="#">Transfer Learning Guide</a>
42	Fine-tuning vs Feature Extraction	<a href="#">Retinal OCT</a>	<i>Deep Learning for Computer Vision</i>
43	Multi-task Learning	<a href="#">CelebA</a>	<a href="#">MTL Survey</a>
44	Mixed Precision Training	ImageNet subset	<a href="#">NVIDIA Apex</a>
45	Projet : Classificateur médical robuste	<a href="#">NIH Chest X-rays</a>	Intégration PyTorch fondations

**Bloc 4 : Computer Vision Médical (Notebooks 46-60)**

#	Titre du Notebook	Dataset	Ressources Clés
46	CNN architectures de base	CIFAR-10	<i>Deep Learning</i> Ch9
47	AlexNet & VGGNet	ImageNet subset	<a href="#">ImageNet Papers</a>
48	ResNet & Skip Connections	<a href="#">ISIC Skin Lesion</a>	<a href="#">ResNet Paper</a>
49	Inception & EfficientNet	<a href="#">Diabetic Retinopathy</a>	<a href="#">EfficientNet Paper</a>
50	Vision Transformers (ViT)	<a href="#">PathMNIST</a>	<a href="#">ViT Paper</a>
51	Data Augmentation avancée (Albumentations)	<a href="#">Blood Cell Images</a>	<a href="#">Albumentations Docs</a>
52	Segmentation avec U-Net	<a href="#">Carvana Image Masking</a>	<a href="#">U-Net Paper</a>
53	Segmentation médicale avancée	<a href="#">LiTS Liver Tumor</a>	<a href="#">nnU-Net</a>
54	Object Detection (YOLO/Faster R-CNN)	<a href="#">RSNA Pneumonia Detection</a>	<a href="#">YOLO Paper</a>
55	GANs pour génération d'images	<a href="#">CelebA</a>	<a href="#">GAN Paper</a>
56	StyleGAN & Medical Imaging	<a href="#">Brain MRI</a>	<a href="#">StyleGAN2 Paper</a>
57	Siamese Networks	<a href="#">Chest X-ray Similarity</a>	<a href="#">Siamese Paper</a>
58	Attention Mechanisms en Vision	<a href="#">Pneumonia X-rays</a>	<a href="#">Attention Paper</a>
59	3D Medical Imaging (3D CNNs)	<a href="#">LUNA16 CT Scans</a>	<a href="#">3D ResNet Paper</a>
60	Projet : Système CAD complet	Multiple medical imaging	Pipeline diagnostique

**Bloc 5 : NLP pour Données Médicales (Notebooks 61-72)**

#	Titre du Notebook	Dataset	Ressources Clés
61	Text Preprocessing médical	<a href="#">MIMIC-III Clinical Notes</a>	<a href="#">spaCy Medical</a>
62	Word Embeddings (Word2Vec/GloVe)	<a href="#">PubMed Abstracts</a>	<a href="#">Word2Vec Paper</a>
63	RNN & LSTM basics	<a href="#">IMDB Reviews</a>	<i>Deep Learning</i> Ch10
64	Bidirectional LSTM & GRU	<a href="#">Medical Transcriptions</a>	<a href="#">Understanding LSTM</a>
65	Attention & Transformer architecture	<a href="#">Medical Q&amp;A</a>	<a href="#">Attention is All You Need</a>
66	BERT fine-tuning médical	<a href="#">Clinical Text Classification</a>	<a href="#">BERT Paper</a>
67	BioBERT & ClinicalBERT	<a href="#">i2b2 Clinical Notes</a>	<a href="#">BioBERT</a>
68	Named Entity Recognition médical	<a href="#">n2c2 NER Dataset</a>	<a href="#">Hugging Face NER</a>
69	Relation Extraction clinique	<a href="#">DDI Extraction</a>	<a href="#">SpERT</a>
70	Question Answering médical	<a href="#">MedQuAD</a>	<a href="#">SQuAD</a>
71	Summarization de notes cliniques	<a href="#">MIMIC-III</a>	<a href="#">BART/T5 Papers</a>
72	Projet : Chatbot médical intelligent	Combinaison datasets NLP	RAG avec LangChain

**Bloc 6 : Séries Temporelles Médicales (Notebooks 73-80)**

#	Titre du Notebook	Dataset	Ressources Clés
73	Time Series basics & ARIMA	<a href="#">ECG Dataset</a>	<i>Forecasting: Principles and Practice</i>
74	LSTM pour séries temporelles	<a href="#">Blood Glucose Levels</a>	<a href="#">LSTM Time Series</a>
75	TCN (Temporal Convolutional Networks)	<a href="#">EEG Signals</a>	<a href="#">TCN Paper</a>
76	WaveNet pour signaux physiologiques	<a href="#">PTB Diagnostic ECG</a>	<a href="#">WaveNet Paper</a>
77	Transformers pour Time Series	<a href="#">Vital Signs Monitoring</a>	<a href="#">Informer Paper</a>
78	Anomaly Detection dans EEG	<a href="#">CHB-MIT Seizure Database</a>	<a href="#">LSTM Autoencoder</a>
79	Multivariate Time Series	<a href="#">MIMIC-III Waveforms</a>	<a href="#">DeepAR</a>
80	Projet : Prédiction événements critiques	<a href="#">eICU Database</a>	Pipeline temps réel

**Bloc 7 : Interprétabilité & Explainability (Notebooks 81-88)**

#	Titre du Notebook	Dataset	Ressources Clés
81	SHAP values	<a href="#">Adult Income</a>	<a href="#">SHAP Documentation</a>
82	LIME pour modèles médicaux	<a href="#">Breast Cancer</a>	<a href="#">LIME Paper</a>
83	Grad-CAM pour images médicales	<a href="#">Chest X-ray</a>	<a href="#">Grad-CAM Paper</a>
84	Integrated Gradients	<a href="#">Skin Lesions</a>	<a href="#">IG Paper</a>
85	Attention Visualization	Modèles NLP précédents	<a href="#">BertViz</a>
86	Counterfactual Explanations	<a href="#">Diabetes</a>	<a href="#">DICE</a>
87	Model Debugging & Error Analysis	Multiple datasets	<i>Interpretable ML</i> (Molnar)
88	Projet : Dashboard d'explainability	Modèles précédents	Streamlit/Dash

**Bloc 8 : Sujets Avancés & Production (Notebooks 89-100)**

#	Titre du Notebook	Dataset	Ressources Clés
89	Federated Learning médical	<a href="#">FEMNIST</a>	<a href="#">Flower Framework</a>
90	Privacy-Preserving ML (Differential Privacy)	Synthetic medical	<a href="#">Opacus</a>
91	Model Quantization & Compression	Modèles précédents	<a href="#">PyTorch Quantization</a>
92	ONNX & Model Deployment	Modèles précédents	<a href="#">ONNX Runtime</a>
93	Docker pour ML Pipeline	N/A	<a href="#">Docker for Data Science</a>
94	MLflow pour Experiment Tracking	Multiple experiments	<a href="#">MLflow Docs</a>
95	FastAPI pour ML Services	Modèles précédents	<a href="#">FastAPI ML Tutorial</a>
96	CI/CD pour ML (GitHub Actions)	N/A	<a href="#">GitHub Actions ML</a>
97	Monitoring modèles en production	Production data	<a href="#">Evidently AI</a>
98	Active Learning pour annotation	<a href="#">Medical Imaging</a>	<a href="#">modAL</a>
99	AutoML avec PyTorch (AutoPyTorch)	<a href="#">OpenML Datasets</a>	<a href="#">Auto-PyTorch</a>
100	Projet Final : Application IA Santé E2E	Datasets personnalisés	Tous les notebooks

---

### Conseils pour Réussir

1. **Ne pas chercher la perfection** : Commence par des versions simples
2. **Documenter abondamment** : Markdown, commentaires, visualisations
3. **Versionner ton code** : Git dès le début
4. **1-2 notebooks par semaine** = programme sur 1 an
5. **Partager sur GitHub** : Portfolio professionnel