Plan

📌 4-Week Plan: Preparation Phase for Master's Thesis (Including Additional Subchapters)

🛠️ **Main Goal:**

- Learn **Lasso, RF, Neural Networks** as core topics.
- Explore additional **ML Engineering & Data Science techniques relevant to medical datasets** (without committing them to the thesis).
- Read and incorporate **relevant subchapters** from *Elements of Statistical Learning (Chapter 3)*.

📆 Week 1 (Feb 10 - Feb 16): Feature Selection & Explainability

📖 Book Chapters to Read:

- **3.1 Introduction** – Overview of regression methods.
- **3.2 Linear Regression Models and Least Squares** – Understanding regression fundamentals.
- **3.2.1 Example: Prostate Cancer** – Case study on medical regression.
- **3.3 Subset Selection** – Methods for feature selection.
- **3.4.2 The Lasso** – L1 regularization and sparsity.

Main Learning (Required)

✅ **Lasso & Group Lasso**: Feature selection using L1 regularization.

✅ **Random Forest (RF) Feature Importance**: Compare with Lasso.

✅ **Handling Missing Data**: Test imputation techniques (SimpleImputer, IterativeImputer).

Side Learning (Optional)

📌 **Explainability Techniques (SHAP, LIME)**

- Study how **SHAP values explain feature importance in medical data**.
- Compare **Lasso-selected features vs. SHAP feature importance**.

📌 **Deliverables**

- Notebook: **Feature Selection (Lasso vs. RF)**
- Notebook (Side Study): **SHAP feature importance study**

📅 Week 2 (Feb 17 - Feb 23): Neural Networks & Hybrid Models

📖 Book Chapters to Read:

- **3.4 Shrinkage Methods** – Ridge Regression (3.4.1) & Lasso (3.4.2).
- **3.4.3 Discussion: Subset Selection, Ridge, and Lasso** – Comparing feature selection & shrinkage.
- **3.8.4 The Grouped Lasso** – Extensions of Lasso.
- **3.8.6 Pathwise Coordinate Optimization** – Efficient Lasso optimization.

Main Learning (Required)

✅ **Neural Networks (NNs) in PyTorch**: Implement a simple NN.

✅ **Hybrid Model**: Train NN to model residuals of Lasso.

✅ **Model Evaluation**: Compare Lasso vs. Lasso + NN performance.

Side Learning (Optional)

📌 **Regularization in Deep Learning**

- Study **dropout, batch normalization, weight decay**.
- Compare **how regularization affects medical dataset predictions**.

📌 **Deliverables**

- Notebook: **Neural Network (PyTorch) + Lasso Residuals**
- Notebook (Side Study): **Regularization experiments**

📅 Week 3 (Feb 24 - Mar 2): ML Engineering & Hyperparameter Tuning

📖 Book Chapters to Read:

- **3.5 Methods Using Derived Input Directions** – PCA & PLS as alternative approaches.
- **3.5.2 Partial Least Squares (PLS)** – Handling correlated features.
- **3.6 Discussion: Comparison of Selection and Shrinkage Methods** – How Lasso compares to Ridge, PCR, and Subset Selection.
- **3.9 Computational Considerations** – Model efficiency and scaling.

Main Learning (Required)

✅ **Hyperparameter Tuning**:

- Use **GridSearchCV/RandomizedSearchCV for RF** (like your friend).
- Use **Bayesian Optimization (Optuna) for NN.**✅ **Comparing Model Performance**: Lasso vs. RF vs. NN.✅ **Medical Dataset Evaluation Metrics**: ROC Curve, Precision-Recall, False Positive Rates.

Side Learning (Optional)

📌 **Model Deployment (Flask, FastAPI)**

- Learn how **ML models are deployed for medical applications**.
- Try **converting a trained model into an API**.

📌 **Deliverables**

- Notebook: **Model Evaluation (Lasso vs. RF vs. NN)**
- Python Script (Side Study): **Flask API with RF Model**

📅 Week 4 (Mar 3 - Mar 9): Writing & Thesis Direction

📖 Book Chapters to Read:

- **3.7 Multiple Outcome Shrinkage and Selection** – Regression for multiple targets.
- **3.8.3 The Dantzig Selector** – Alternative sparse estimation method.
- **3.8.5 Further Properties of the Lasso** – Understanding Lasso behavior.
- **3.9 Computational Considerations** – Optimizing regression models for efficiency.

Main Learning (Required)

✅ **Summarize experiments & findings.**

✅ **Decide on Master's Thesis focus (ML Engineering vs. Data Science).**

✅ **Write the 3-5 page report & work plan for thesis.**

Side Learning (Optional)

📌 **Optimization & Efficiency in ML**

- Study **model quantization & pruning** for optimizing medical models.
- Understand **how to make ML models run efficiently in real-world medical applications**.

📌 **Deliverables**

- **Final Report (3-5 pages).**
- Notebook (Side Study): **Pruned NN Model Experiment**.

📌 Summary: Core Work vs. Side Learning vs. Book Chapters

Week	Main Learning (Required)	Side Learning (Optional, Just Studying)	Book Chapters
Week 1	Lasso, Group Lasso, RF Feature Selection	SHAP & LIME (Explainability)	**3.1, 3.2, 3.3, 3.4.2, 3.2.1 (Prostate Cancer Example)**
Week 2	Neural Networks (PyTorch) + Lasso Residuals	Regularization (Dropout, Weight Decay)	**3.4, 3.8.4, 3.8.6, 3.4.3 (Subset vs. Ridge vs. Lasso)**
Week 3	Hyperparameter Tuning (GridSearch, Optuna)	Flask/FastAPI Model Deployment	**3.5, 3.6, 3.9, 3.5.2 (PLS), 3.6 (Comparison of Shrinkage Methods)**
Week 4	Final Report & Thesis Decision	Model Optimization (Quantization, Pruning)	**3.7, 3.8.3, 3.8.5 (Lasso Properties), 3.9 (Computational Considerations)**
