

## **Minimal Updated Paper Blueprint (Based on Final Code)**

### **Research Question**

**Which machine learning method best predicts next-day SPY price direction, and does any model outperform logistic regression and a naive benchmark?**

#### **1. Abstract (½ page)**

- Objective: predict next-day stock price direction
- Data: SPY daily data (2015–2025)
- Models: logistic regression, LASSO (LARS + glmnet), GAM, decision tree, random forest
- Evaluation: AUC, accuracy, F1, deviance
- Main result: no model materially outperforms the naive benchmark or logistic regression

#### **2. Introduction (≈1 page)**

- Motivation: difficulty of short-horizon stock prediction
- Importance of testing ML vs traditional models
- Research question
- Brief overview of approach and structure

#### **3. Data & Feature Construction (≈1 page)**

- Data source: tidyquant (SPY OHLCV)
- Sample period: 2015–2025
- Target variable:
  - UpTomorrow = 1 if next-day log return > 0
- Feature groups used:
  - Lagged log returns (1, 2, 10, 30)
  - Normalized trend indicators (distance to SMA)
  - Volatility (10-day rolling)
  - Momentum (10, 30)
  - Normalized daily range and open–close move
  - Relative (log) volume
- Note on normalization due to price non-stationarity

#### **4. Methodology (≈2 pages)**

##### **4.1 Train/Test Split**

- Chronological split (no shuffling)
- Main split: 85% train / 15% test
- (remember to do the 70/30 split if we have space)

##### **4.2 Models Estimated**

- Logistic regression (baseline)

- LASSO:
  - LARS for variable selection + logit refit
  - Penalized logit via glmnet
- GAM with nonlinear splines
- Decision tree
- Random forest
- Naive benchmark (unconditional up-day probability)

### **4.3 Evaluation Metrics**

- Accuracy
- Precision, recall, F1
- ROC curve and AUC
- Normalized deviance (log-loss)

### **5. Results (≈2 pages)**

- Main performance comparison table (all models)
- ROC curve(s)
- Random forest variable importance
- Key finding:
  - Most models perform close to random
  - LASSO selects intercept-only model
  - Naive benchmark matches or exceeds ML models

### **6. Discussion (≈1–1.5 pages)**

- Interpretation of weak predictive performance
- Comparison across model classes (linear vs nonlinear vs tree-based)
- Implications for market efficiency
- Limitations:
  - Single asset
  - Daily horizon
  - No trading strategy evaluation

### **7. Conclusion (½ page)**

- Answer research question directly
- Summary of findings
- Suggest future extensions (longer horizons, more assets)

**Appendix:** Full R code, Additional figures and tables