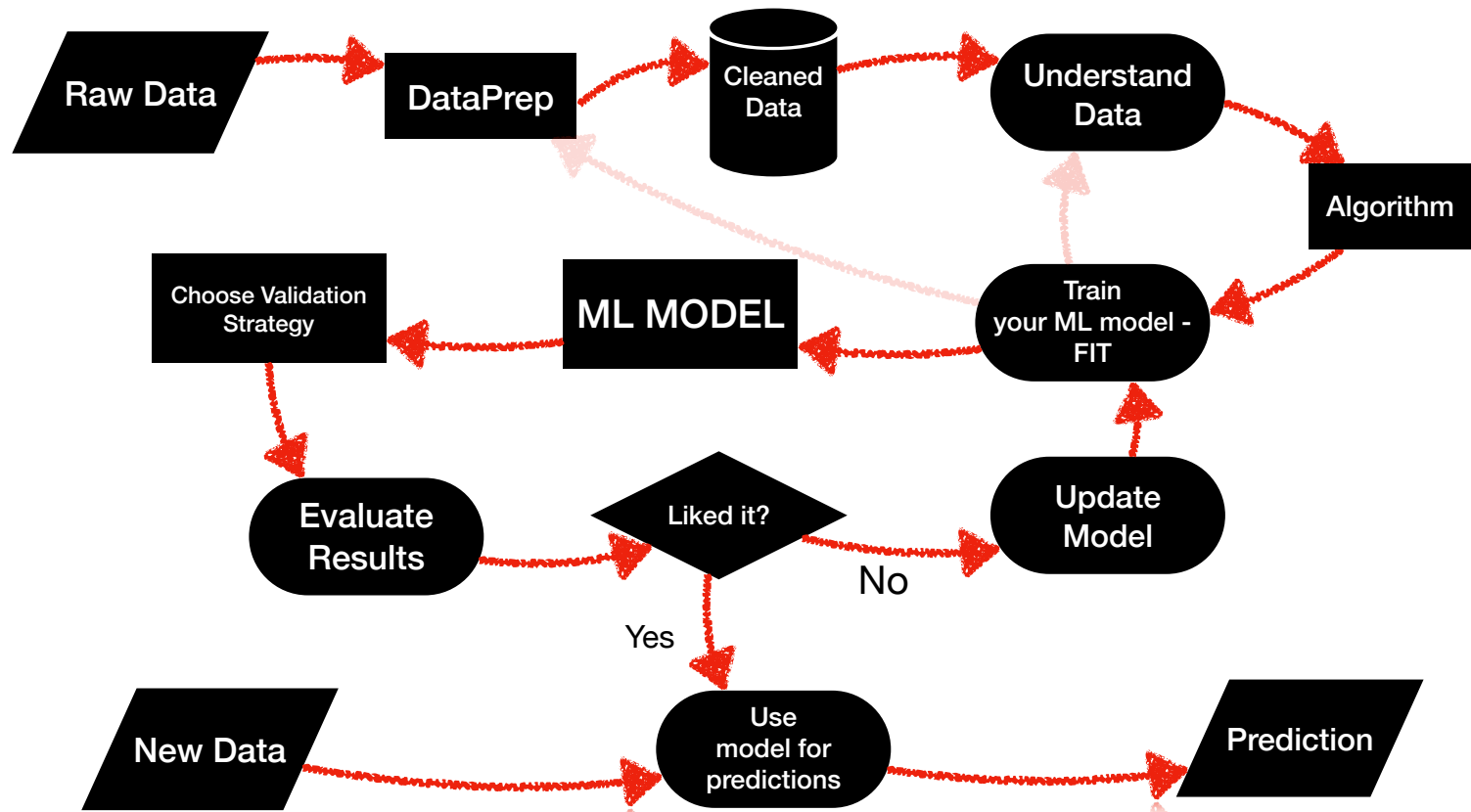
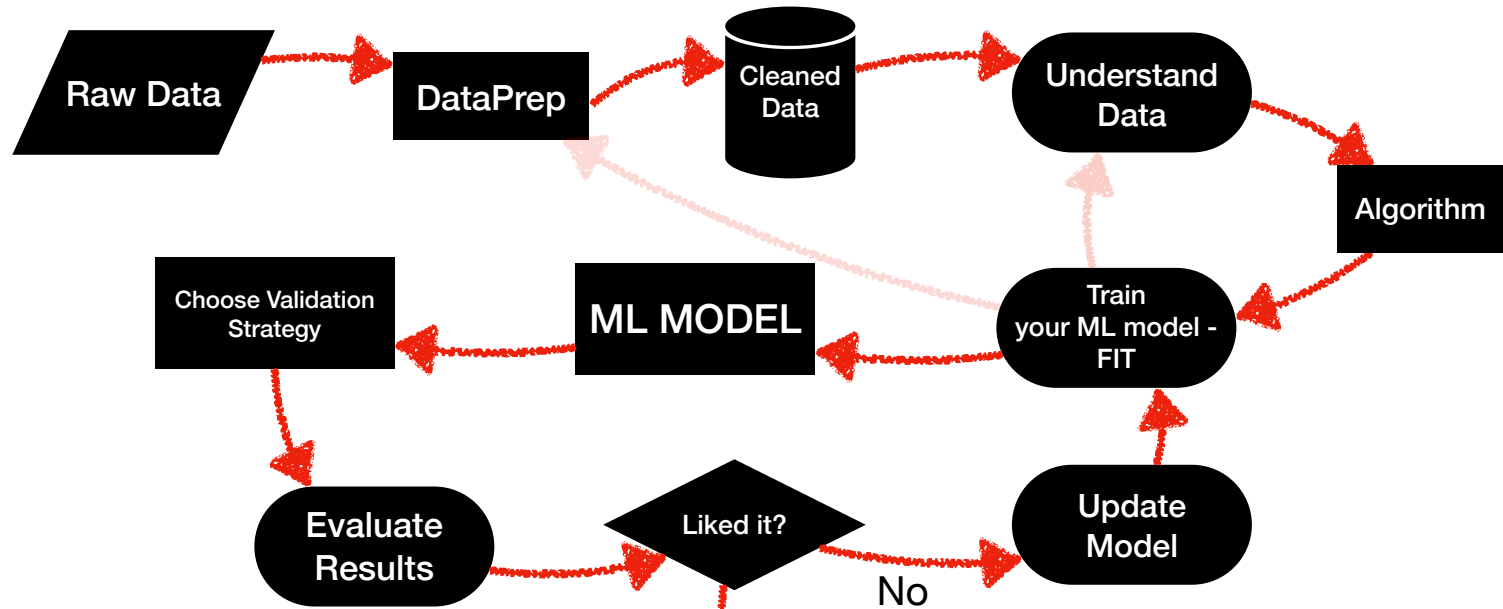




## Machine Learning Workflow



## Development



## Production or Deploy





## Overfitting

- The model has memorized the data
- It could not extract the **general** patterns
- The model haven't **generalized** well enough.



## Training Serving Skew

- **Backtests**
- Performance in backtests are fine
- Performance in deployment is poor
- Usually caused because data is treated differently in training vs in production



## Concept Drift

- **Monitoring**
- Patterns can change over time
- Your model deteriorates over time.
- Your model should be retrained



## Overfitting

- Model too complex
  - Regularization
  - Feature Selection
- Data leakage
- Validation strategy



## Training Serving Skew

- **Backtests**
- Performance in backtests are fine
- Performance in deployment is poor
- Usually caused because data is treated differently in training vs in production



## Concept Drift

- **Monitoring**
- Patterns can change over time
- Your model deteriorates over time.
- Your model should be retrained

