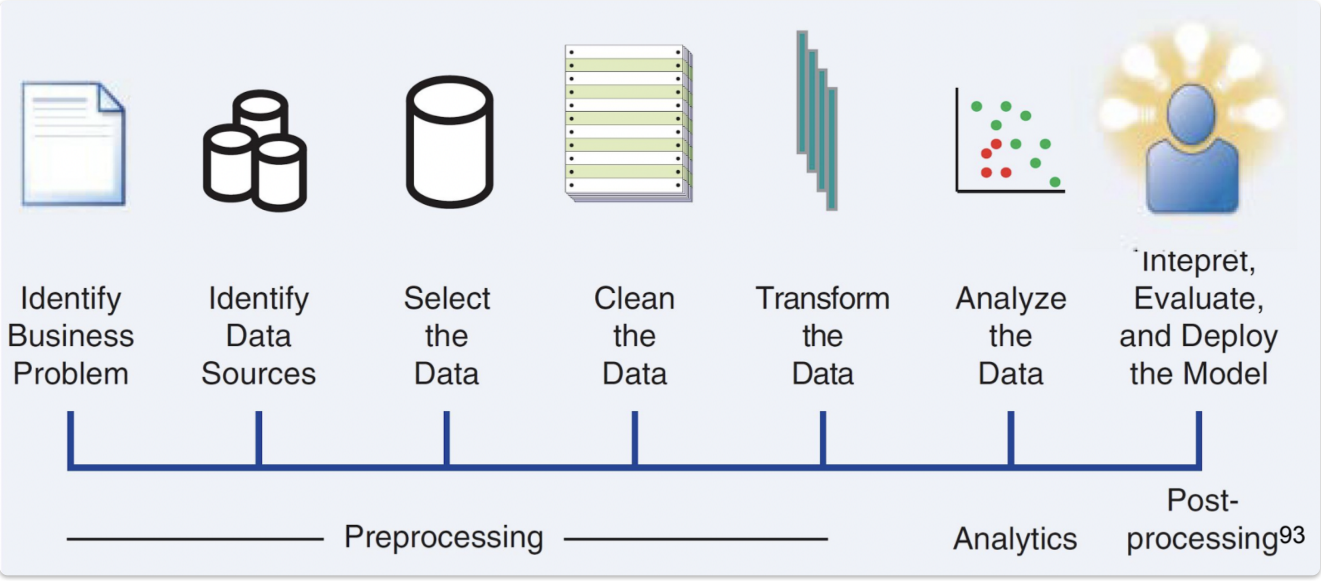
2.5. Fraud Analytical Process #AnalyticalProcess

2.5.1 The Fraud Analytics Process Model



1. Identify Data Sources:
Data are the key ingredient to any analytical exercises.

2. Select the Data:
Data selection has an **impact** on the *analytical models*:

- Data gathered in a staging area.
- Basic exploratory analysis.

3. Clean the Data:
Get rid of all inconsistencies, such as missing values and duplicate data.

4. Transform the Data:
Additional transformations: binning, alphanumeric to numeric coding, geographical aggregation, and so on.

5. Analyze the Data:
The analytical model is estimated on the *preprocessed and transformed data*.
The actual fraud-detection model is built.

6. Interpret, Evaluate and Deploy the Model:
The *model is interpreted and evaluated* by the fraud experts.

2.5.1.1 Possible Analysis Output

- **Trivial fraudulent patterns**: *Validation of the model*.

- **Unknown patterns**: provide *added insight and detection power* ("knowledge diamonds").

Once the analytical model has been appropriately validated and approved, it can be put into *production*.

2.5.1.2 Additional Considerations

- How to represent the model output in a user-friendly way?
- How to integrate it with other applications?
- How to make sure the analytical model is appropriately monitored and back tested on an ongoing basis?

2.5.2 Key characteristics of successful fraud analytics models

A fraud-detection model must be thoroughly evaluated before being adopted.

🔑 Key characteristics of successful fraud analytics models:

- Statistical accuracy.
- Interpretability.
- Operational efficiency.
- Economical cost.
- Regulatory compliance.

2.5.2.1 Statistical accuracy #StatisticalAccuracy

#DEF **Detection power and correctness of the statistical model in flagging suspicious cases.**

- Different Metrics

> *We need to make sure that the model generalizes well and is not overfitted to the historical data set.*

2.5.2.2 Interpretability #Interpretability

#DEF **When a deeper understanding of the detected frauds is required, a fraud-detection model must be *interpretable*.**

Model's interpretability depends on the technique used.

- **White-box models**: *Allow to understand* the underlying reasons why the model signals a case to be suspicious.
- **Black-box models**: Complex, *non interpretable models*.

2.5.2.3 Operational efficiency #OperationalEfficiency

#DEF **Time** and **effort** that is required to:

- Collect and preprocess the *data*.
- To *evaluate the model*.
- *Monitor and backtest* the model, and re-estimate it.
- To *evaluate* whether *a case* is suspicious or not.

⚠ When **cases need to be evaluated in real time** - > **operational efficiency is crucial** and is a **main concern during model performance assessment**.

2.5.2.4 Economical cost #EconomicalCost

#DEF **Developing and implementing a fraud-detection model involves a significant cost to an organization**:

- To gather, preprocess, and analyze the *data*.
- To put the *resulting analytical models into production*.
- The *software, human, and computing resources*.
- *External data* to enrich the available in-house data.

> ***Cost-benefit analysis*** *to gain insight in the constituent factors of the returns on investment of building an advanced fraud-detection system.*

Security vs. Cost Balance

1. *Direct costs*:
    - Management.
    - Operational.
    - Equipment.
2. *Indirect costs (more relevant)*:
    - Less usability.
    - Slower performance.
    - Less privacy (due to security controls).
    - Reduced productivity (users are slower).

⚠**More money =/=> More security**:

- Very expensive, “unconfigured” Fraud Detection System:
    - Better not to have it.
- Complex authentication that slows down users:
    - Users will write passwords on stickies.
- etc...

2.5.2.5 Regulatory compliance #RegulatoryCompliance

#DEF **A fraud-detection model should be in line and comply with all applicable regulation and legislation.** (e.g., PSD2)

> *Depending on the context there may be internal or organization-specific and external regulation that applies to the development and application of a model.*

2.5.3 Challenges of developing fraud-detection models

Challenges:

- Dynamic nature of fraud.
- Accuracy.
- Skewness of the data.
- Operational Efficiency.
- Evaluation time and Big Data management.

2.5.3.1 Dynamic nature of fraud #NatureOfFraud

Fraudsters constantly try to beat detection and prevention systems by developing **new strategies and methods**.

> *Adaptive analytical models for detection and prevention systems are **required**, in order to detect and resolve fraud as soon as possible.*

2.5.3.2 Accuracy #Accuracy

- **Good detection power**: Detect fraud as *accurately as possible*.
- **Not to miss out on too many fraud cases**, especially *involving a large amount or financial impact*.
- **Low false alarm rate**, to avoid harassing good customers and prevent accounts or transactions to be blocked unnecessarily.

The cost of missing a fraudulent case may be significant.

2.5.3.3 Skewness of the data #Skewness

#DEF **Skewness**: We typically have *plenty of* historical examples of *non-fraudulent cases*, but only a *limited number of fraudulent cases*.

⚠ **Needle-in-a-haystack problem** -> might cause an analytical technique to experience difficulties in learning an accurate model.

2.5.3.4 Operational Efficiency #OperationalEfficiency

#DEF **Limited amount of time available to reach a decision and let a transaction pass or not.**

Such a requirement clearly *impacts*:

- The *design* of the **operational IT systems**.
- The *design* of the **analytical model**.

2.5.3.5 Evaluation time and Big Data management #EvaluationTime

#DEF **Must be able to deal with the massive volumes of data that are available and need to be processed.**

- Must be *able to deal with the massive volumes of data* that are available and need to be processed.
- The *information or the variables* that are used by the model *should not take too long to be gathered or calculated*.

Next chapter: Red Flags of Frauds