# Automated Patent Classification

## What we have learned from client projects

Evalueserve
IP and R&D Solutions

# Agenda

- Introduction

- Automated Patent Classification @ Evalueserve IP and R&D Projects

- Use Cases & Key Findings

- Appendix

# Presenters

## Jeroen Kleinhoven

Global Head of Product Strategy
**IP and R&D Business Unit**

- Director Evalueserve Netherlands
- The Hague, The Netherlands
- Joined Evalueserve in April 2015
- Former CEO Treparel
- Bachelor Small Business Management (Haarlem)
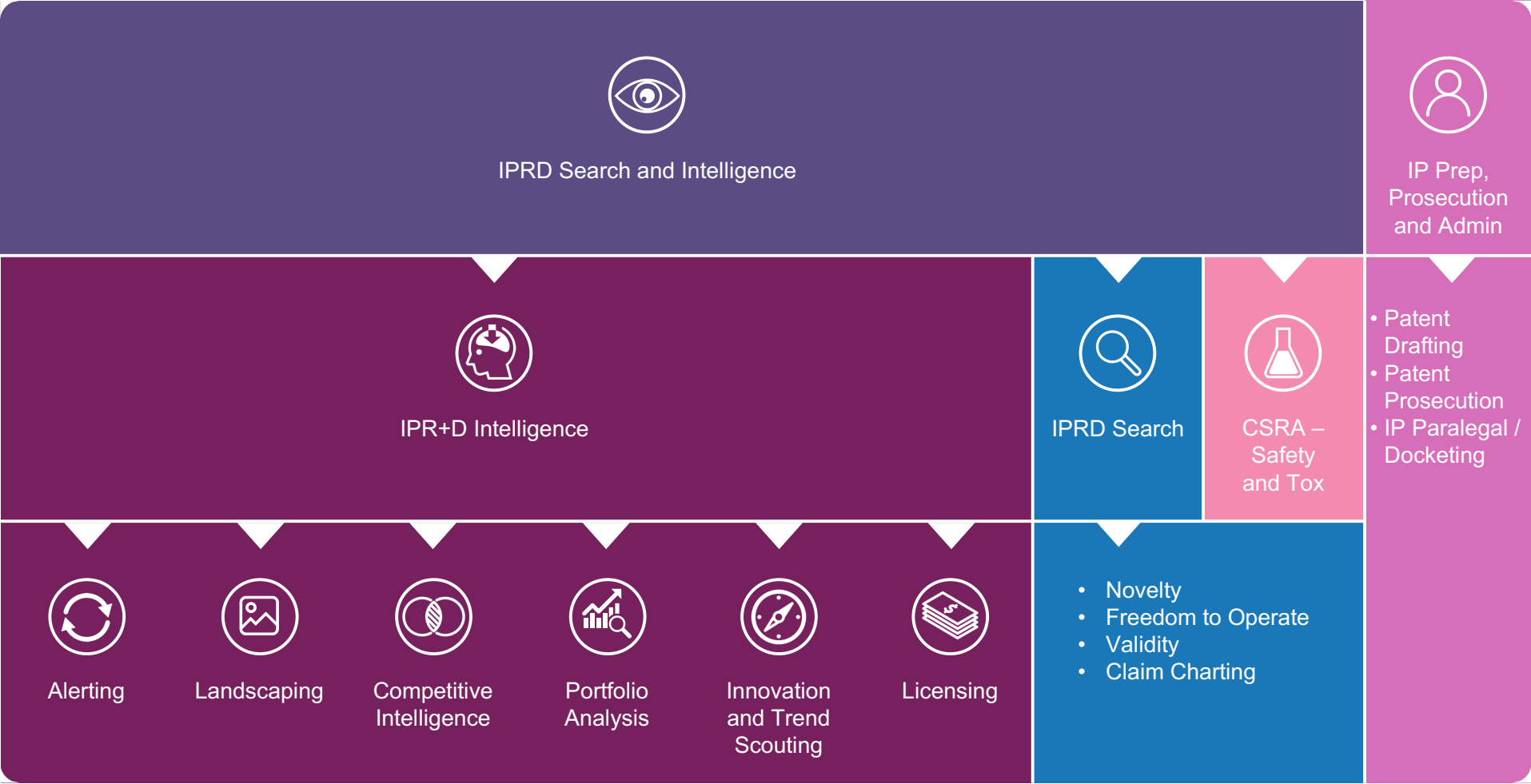
## Dr. Fedde van der Lijn

Solution Engineer
**IP and R&D Business Unit**

- Global Competence Centre for Text Mining IP R&D
- The Hague, The Netherlands
- Joined Evalueserve in April 2015
- Former Head of Solutions at Treparel
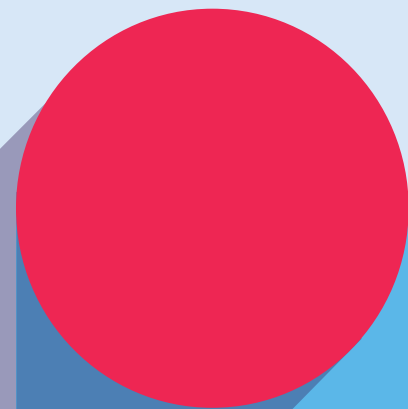- PhD in Biomedical Image Analysis (Rotterdam)

- Treparel developed KMX Patent Analytics in close cooperation with Royal Dutch Philips (2007)
- Evalueserve is a global professional services provider offering research, analytics and data management services.
- With over 3,200 professionals worldwide we are powered by mind+machine™ – a unique combination of human expertise and best-in-class technologies.

# Our solutions for Global Leaders in IP and R&D

IPRD Search and Intelligence

IP Prep, Prosecution and Admin

IPR+D Intelligence

IPRD Search

CSRA – Safety and Tox

- Patent Drafting
- Patent Prosecution
- IP Paralegal / Docketing

Alerting

Landscaping

Competitive Intelligence

Portfolio Analysis

Innovation and Trend Scouting

Licensing

- Novelty
- Freedom to Operate
- Validity
- Claim Charting

# Automated Patent Classification @ Evalueserve IPR+D

# Typical Evalueserve IPR+D project workflow

Evalueserve IPR+D Intelligence Solution: i.e. Patent Landscape

**1**

### Search and Import

Search and import potentially relevant documents from 3rd party data sources

**2**

### Filter and Tag

Review and identify relevant and non-relevant documents.

Label documents to the Evalueserve Tagging framework

**3**

### Client Review

Clients have option to review filtering and tagging results

Review workflows can be adjusted to client operations

**4**

### Report & Deliver

Create online reports and share project with end users

Single client repository for all IP and R&D data: relevant and tagged

# Filtering and tagging

I.e. patent landscape on OLED technology

Comparable to preclassification

Comparable to reclassification

US9287331B2 → **Filtering** → US9287528B2 → **Tagging** → US9287528B2

tag1 tag2
tag3 tag4

**Broad result set**

**(based on keyword, IPC queries)**

**Narrow result set**

**Tagged result set**

**KMX Classifiers**

| | | |
|---|---|---|
| ▼ | Application | |
| | ▼ | Display Utilization |
| | | Not Described |
| ▼ | Problems | |
| | ▼ | Benefit |
| | | Fast Response |
| | ▼ | Technical Challenges |
| | | Power Management |
| ▼ | Solutions | |
| | ▼ | Device |
| | | ▼ Back Panel |
| | | ▼ Pixel Circuit |
| | | Switching Circuit |

# How do KMX classifiers work?

- KMX classifiers are based on a word fingerprint (~1000 most prominent terms in broad result set)
- Classifiers need to be trained using manually tagged examples
  - Filtering: relevant / non-relevant
  - Tagging: i.e. Fast Response / Lifespan / Low Power / …
- Obtain training data by:
  - Tagging a subset of the broad result set
  - Using previous analyses
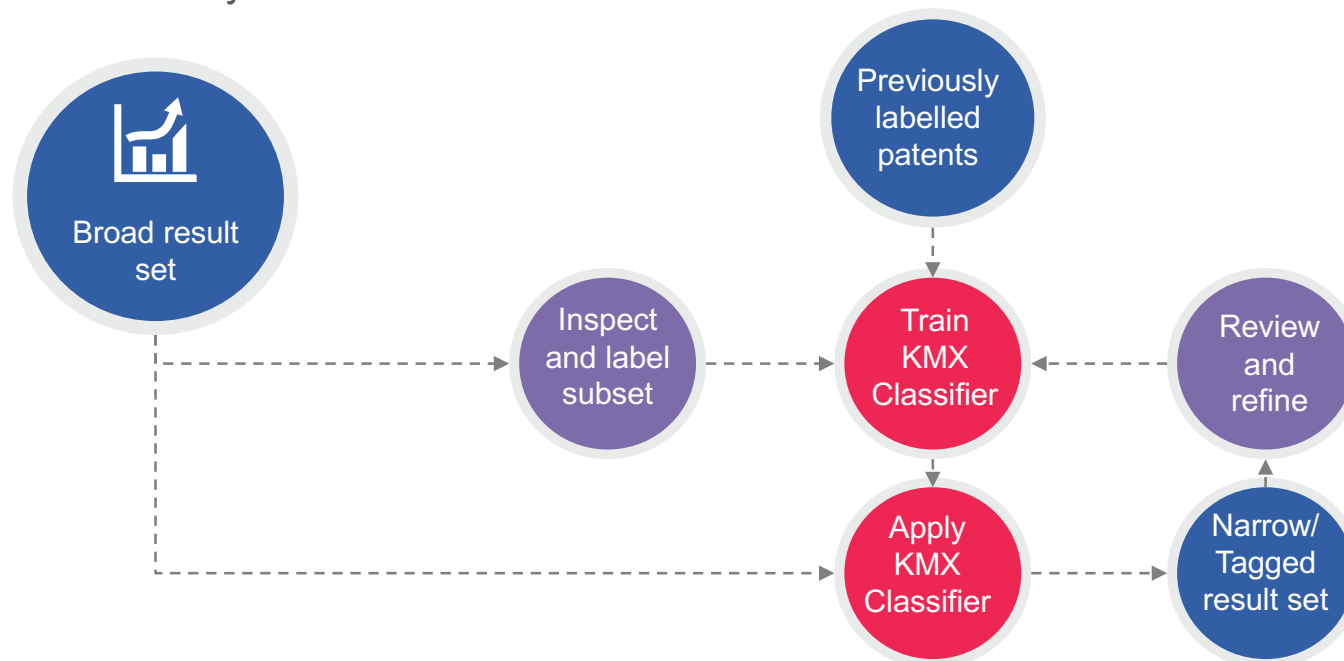
# How do KMX classifiers work?

- After training classifiers can run automatically

  - Option to review and refine the training set to improve results

- KMX classifiers generate tag probabilities

  - Filtering: i.e. 95% relevant / 5% non-relevant

  - Tagging: i.e. 67% Fast Response / 12% Lifespan / 15% Low Power / 6% …
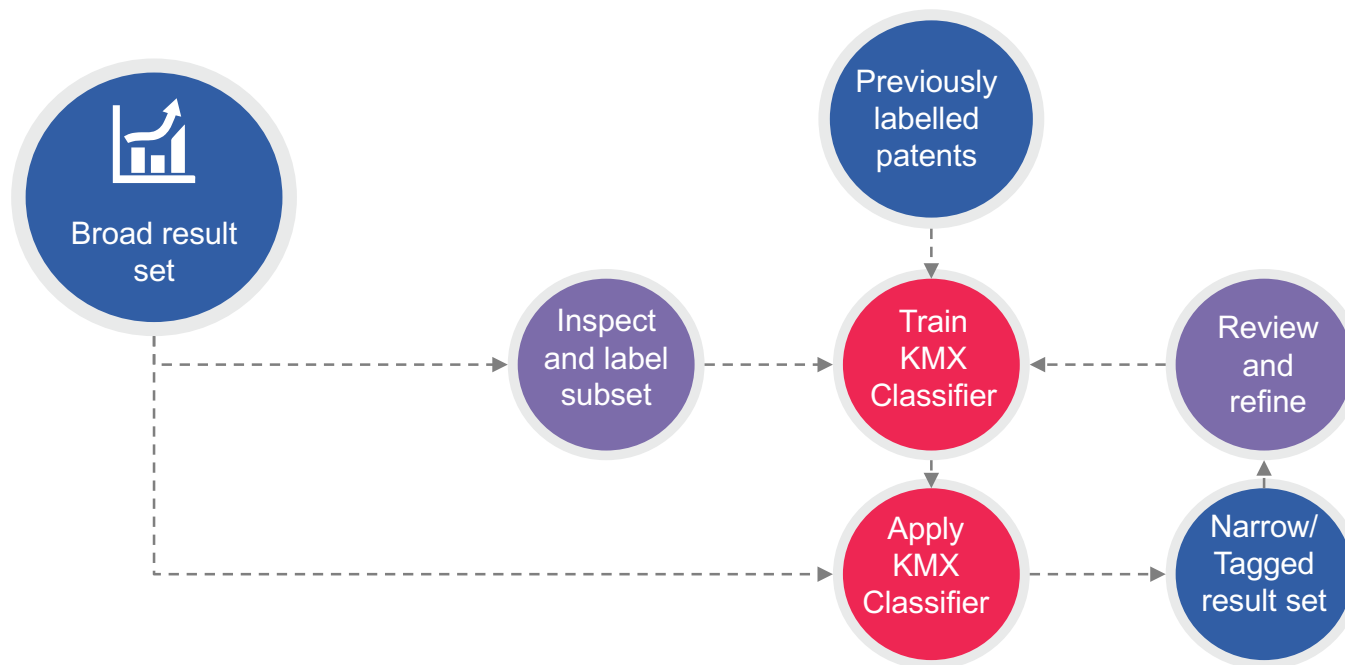
# How does Evalueserve IPR+D use KMX classifiers?

- Specific workflow depends on client requirements:

  - Mind + Machine workflows

    - **Accuracy** is key requirement

    - Possible workflows

      - Classifiers provide first tags, expert reviews

      - Classifiers and expert both provide tags, expert reviews discrepancies

      - Example use cases can be found in appendix

  - Machine only workflows
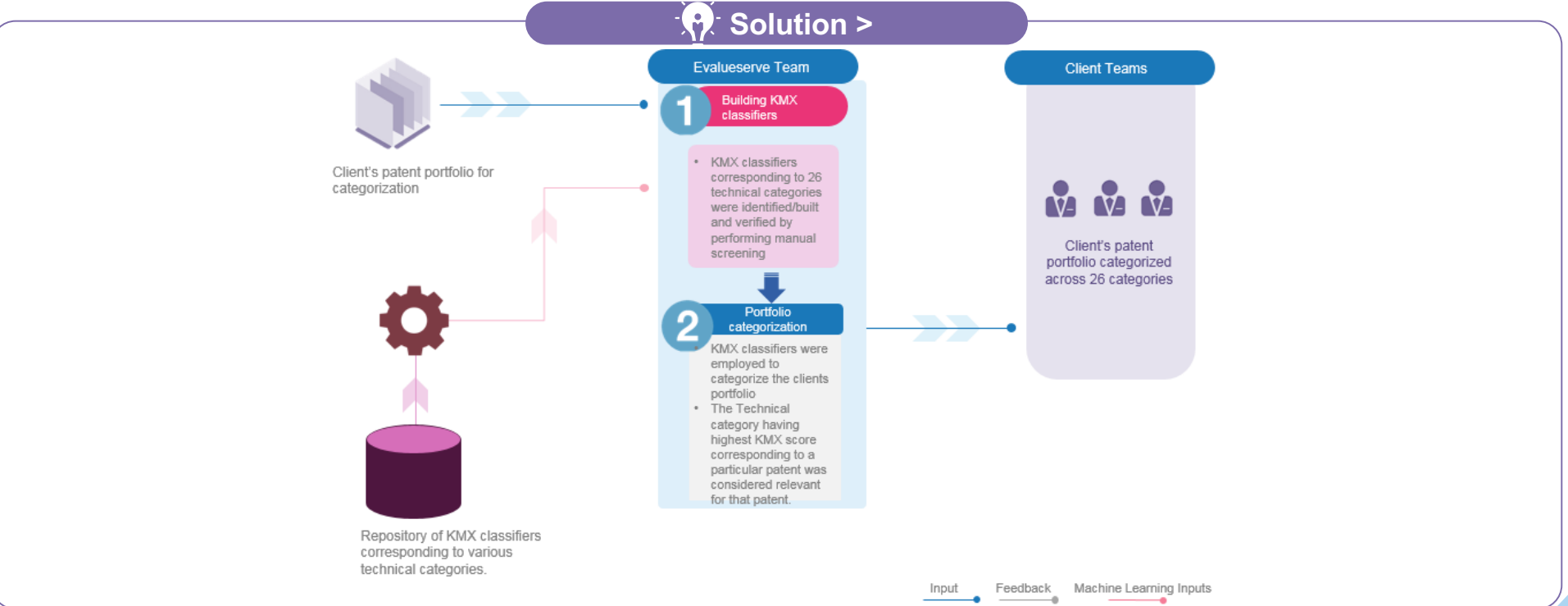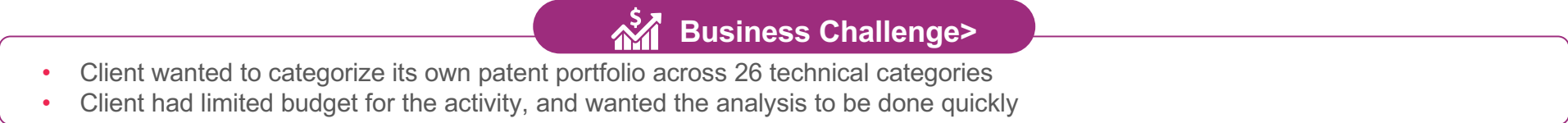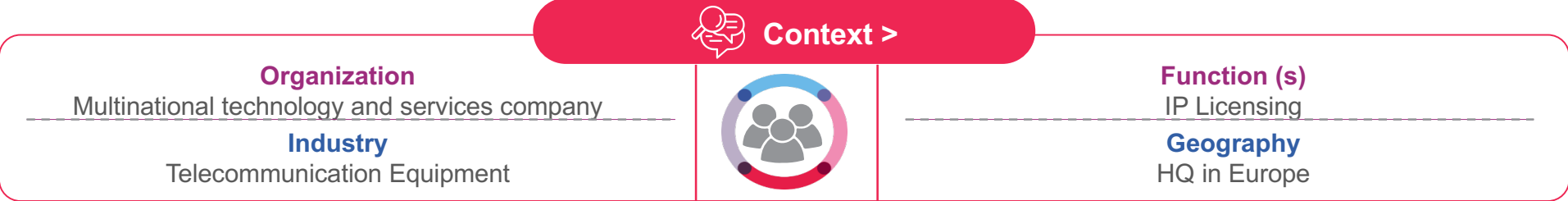
    - **Efficiency** is key requirement

# Machine-only use cases
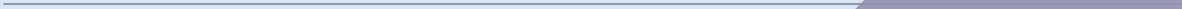
# Case Study 1 – Portfolio Categorization
## Automatically assign technology tags to patent portfolio

### Context >

**Organization**
Multinational technology and services company

**Industry**
Telecommunication Equipment

**Function (s)**
IP Licensing

**Geography**
HQ in Europe

### Business Challenge>

- Client wanted to categorize its own patent portfolio across 26 technical categories
- Client had limited budget for the activity, and wanted the analysis to be done quickly

### Solution >

Client's patent portfolio for categorization

Repository of KMX classifiers corresponding to various technical categories.

**Evalueserve Team**

**1** Building KMX classifiers

- KMX classifiers corresponding to 26 technical categories were identified/built and verified by performing manual screening

**2** Portfolio categorization

- KMX classifiers were employed to categorize the clients portfolio
- The Technical category having highest KMX score corresponding to a particular patent was considered relevant for that patent.

**Client Teams**

Client's patent portfolio categorized across 26 categories

Input        Feedback        Machine Learning Inputs
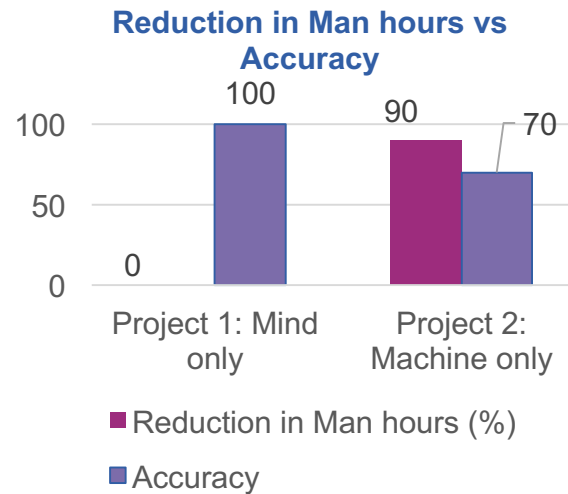
# Case Study 1 – Portfolio Categorization
## Achieved 90% Time savings with Machine approach

### Approach >

**Steps**:

**"Machine" steps -**

- Use already manually categorized patents to create KMX classifiers

- Test each classifier with sample set of patents to fine-tune classifiers.

- Use KMX classifier to assign 'score' to each patent to obtain automated categorization

**Reduction in Man hours vs Accuracy**

| | Project 1: Mind only | Project 2: Machine only |
|---|---|---|
| Reduction in Man hours (%) | 0 | 90 |
| Accuracy | 100 | 70 |

■ Reduction in Man hours (%)
■ Accuracy

### Benefit >

**Productivity**

- **90% reduced cost for client**

**Quality**

- **Achieved 70% accuracy**

# Case Study 2 – Proactive Patent Defence
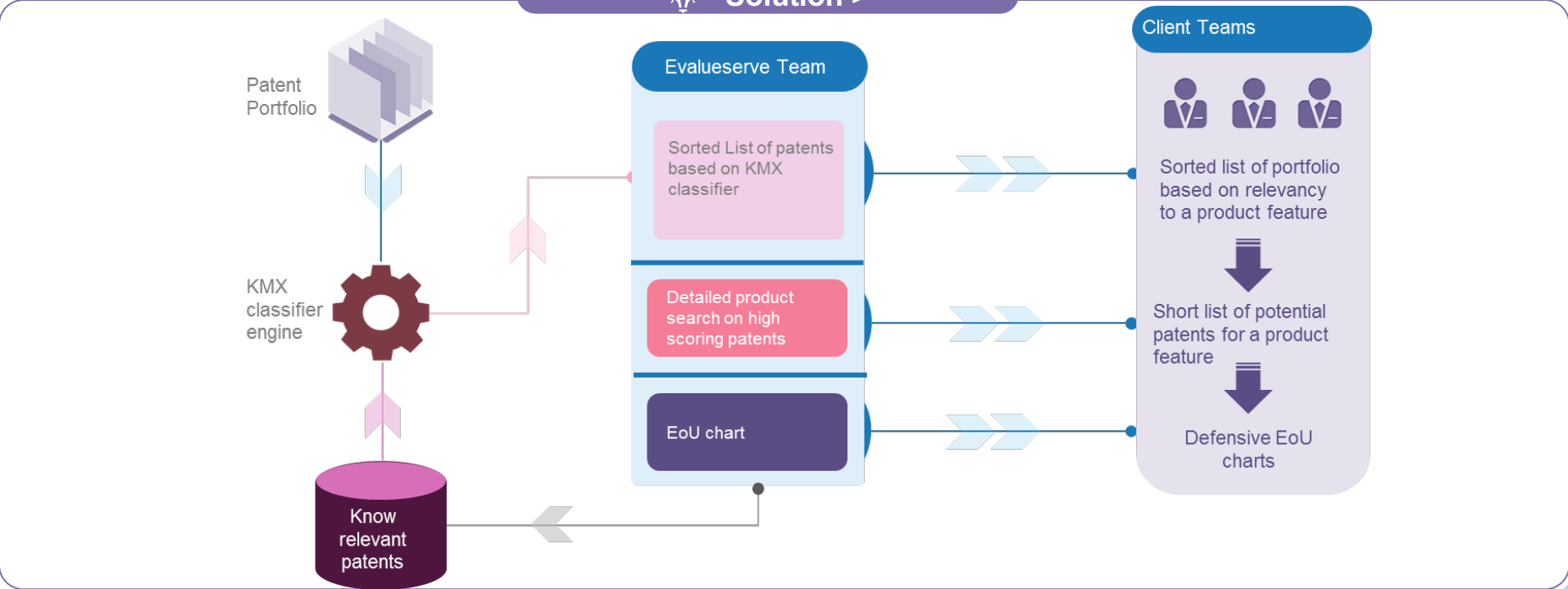
Identifying defensive patents against companies posing threat

## Context >

**Organization**
Multinational technology corporation

**Industry**
Internet, Web services

**Function (s)**
Due Diligence

**Geography**
HQ in US

## Business Challenge >

- Client wants to be quickly ready with defensive list of patents against companies posing threat to them
- Client wants to identify list of patent from their own portfolio which can be relevant to features of products from company posing threat

## Solution >

Patent Portfolio

KMX classifier engine

Know relevant patents

**Evalueserve Team**

Sorted List of patents based on KMX classifier

Detailed product search on high scoring patents

EoU chart

**Client Teams**

Sorted list of portfolio based on relevancy to a product feature

Short list of potential patents for a product feature

Defensive EoU charts

# Case Study 2 – Proactive Patent Defence
## Reduced 35% effort to find a set of relevant patents

### Approach >

**Steps**:

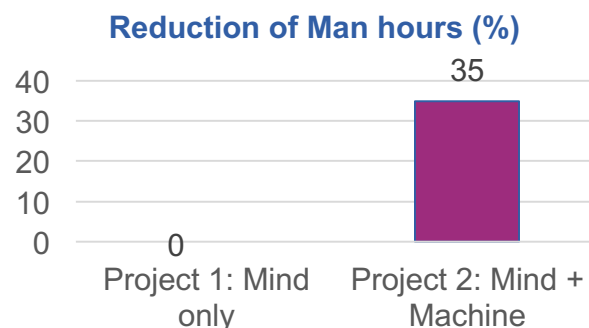**"Machine" steps -**

- Identify training set for KMX classifier based on known patents relevant for a product feature
- Use various parameter weight setting to identify optimal setting based on KMX landscape
- Identify negative training set from KMX landscape
- Build KMX classifier using the training set and run it on companies own portfolio

**"Mind" step, done by client himself -**

- Analyze top patents from KMX classifier

**Reduction of Man hours (%)**

| | |
|---|---|
| Project 1: Mind only | 0 |
| Project 2: Mind + Machine | 35 |

**Client obtained a sufficient number of patents in both projects**

### Benefit >

**Productivity**

- **35% reduced effort of client**

**Quality**

- **Client obtained a sufficient number of patents for the defensive list from relevant patents identified using KMX**

# Key takeaways

- Automated classification is used on a regular basis by Evalueserve
    - To improve accuracy: mind + machine approach
    - To increase efficiency: machine-only approach

- Proven use cases:
    - It can be used for:
        - Filtering relevant documents out of a broad result set (comparable to pre-classification)
        - Tagging documents to a hierarchical label tree (comparable to reclassification)

- Case examples
    - Automated classification can lead to clear efficiency gains
        - Portfolio Categorization use case:
            - Efficiency gain: 90%
            - Accuracy: 70%
        - Proactive Patent Defense use case:
            - Efficiency gain 30%
            - Accuracy: 100%*

*Client obtained sufficient number of relevant patents

# Thank you!



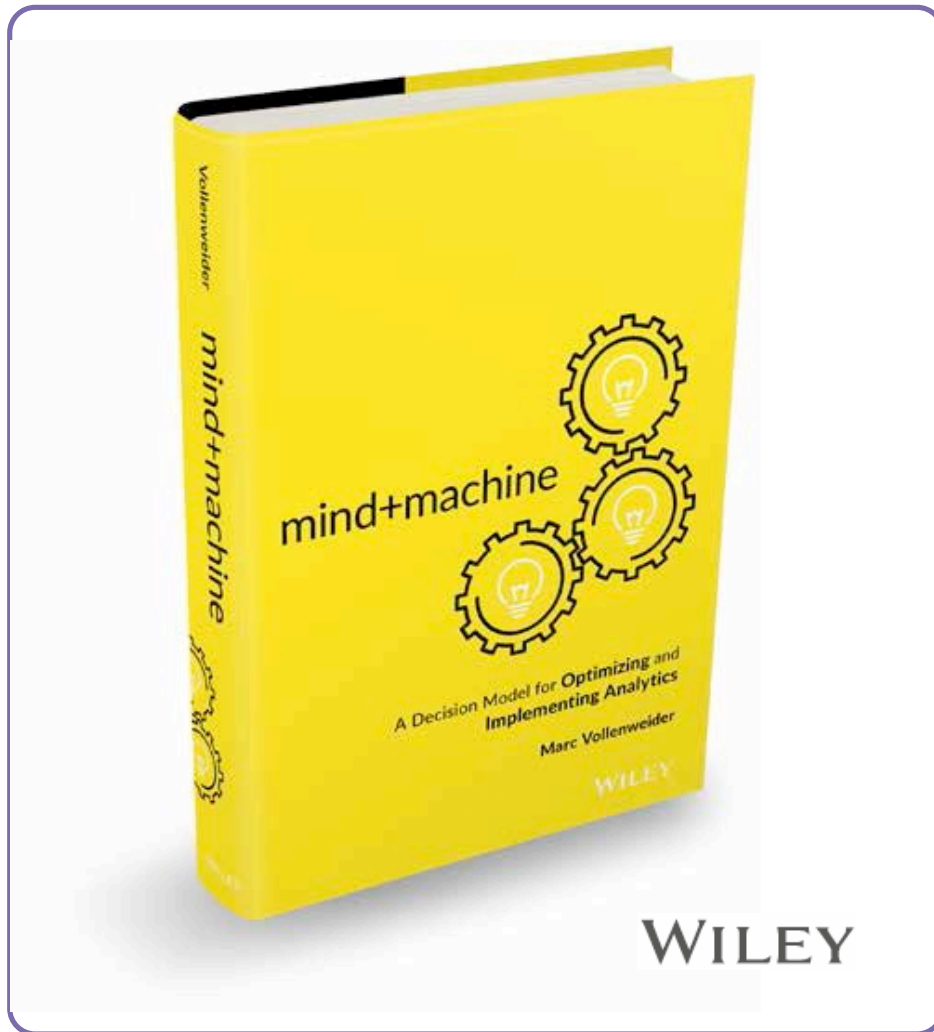Jeroen Kleinhoven
(jeroen.kleinhoven@evalueserve.com)



Fedde van der Lijn
(fedde.vanderlijn@evalueserve.com)

# Recommended read: mind+machine
a decision model for optimizing and implementing analytics



**By Evalueserve Co-Founder,
Chief Strategist Marc Vollenweider**

Decision-makers are reliant on receiving the necessary insight at the right time in a suitable format.

The marriage of the perceptive power of the human brain with the benefits of cutting edge machines is essential because neither mind nor machine can handle the complexities of modern analytics on their own.

Only when the two come together with structure and purpose to solve a problem are goals achieved.

This book by Marc Vollenweider covers the entire end to-end value chain of analytics.
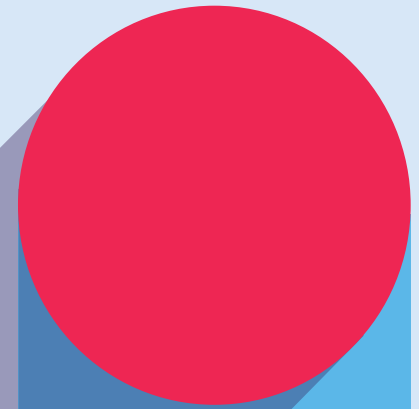
**Mind+Machine: A Decision Model for Optimizing and Implementing Analytics**
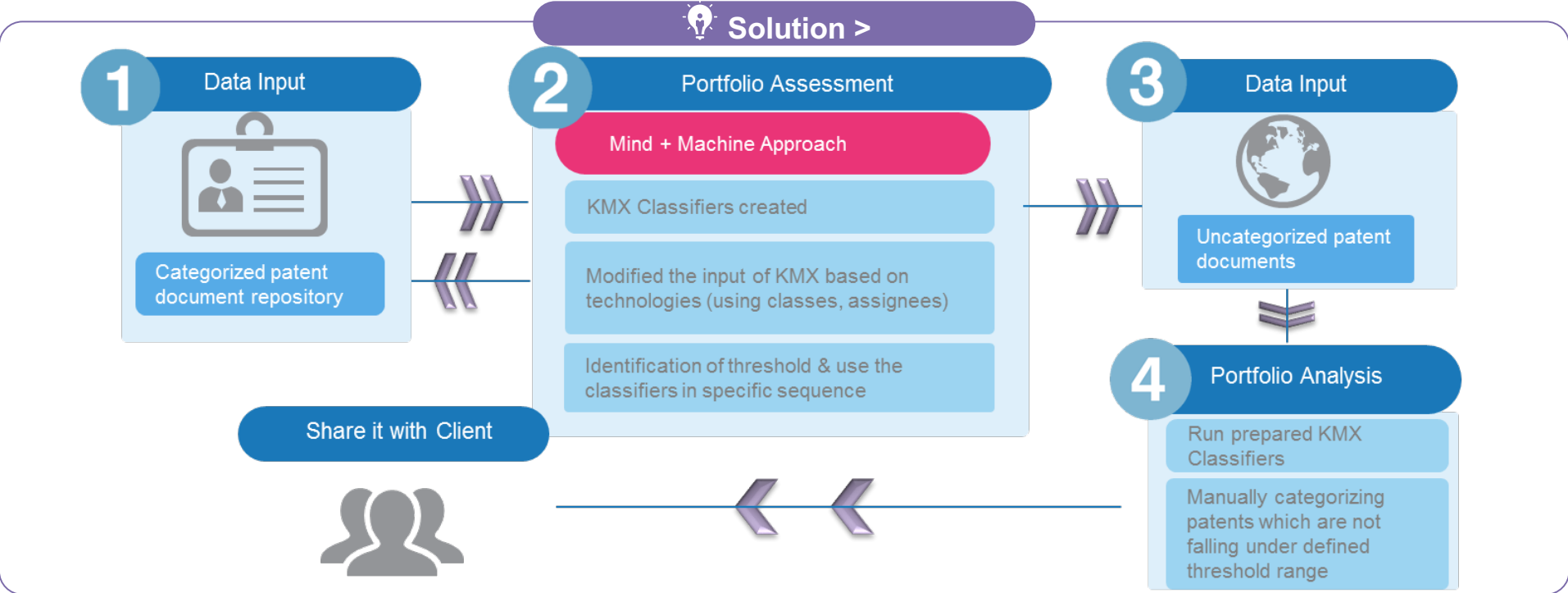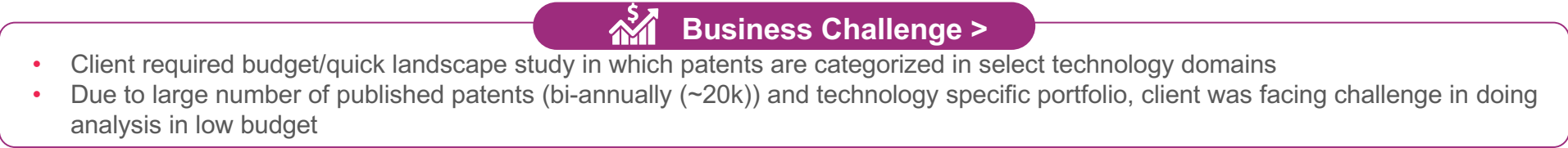by Marc Vollenweider
**Link:** http://a.co/1eZyrBs

# Appendix – Mind+Machine use cases

# Case Study 3 – Managing bi-annually portfolio updates
## Implementing Mind+Machine for identification of domain of patent documents

**Context >**

**Organization**
Multinational Conglomerate

**Industry**
Electric, Aviation, Healthcare, Oil and Gas

**Function (s)**
IP Strategy

**Geography**
HQ in US

**Business Challenge >**

- Client required budget/quick landscape study in which patents are categorized in select technology domains
- Due to large number of published patents (bi-annually (~20k)) and technology specific portfolio, client was facing challenge in doing analysis in low budget

**Solution >**

**1** Data Input

Categorized patent document repository

**2** Portfolio Assessment

Mind + Machine Approach

KMX Classifiers created

Modified the input of KMX based on technologies (using classes, assignees)

Identification of threshold & use the classifiers in specific sequence

Share it with Client

**3** Data Input

Uncategorized patent documents

**4** Portfolio Analysis

Run prepared KMX Classifiers

Manually categorizing patents which are not falling under defined threshold range

# Case Study 3 – Managing bi-annually portfolio updates

Achieved 90% accuracy with Mind+Machine approach, with time savings

## Approach >

**Steps**:

**"Machine" steps -**

- Use already manually categorized patents to create KMX binary classifiers
- Test each classifier with sample set of patents.
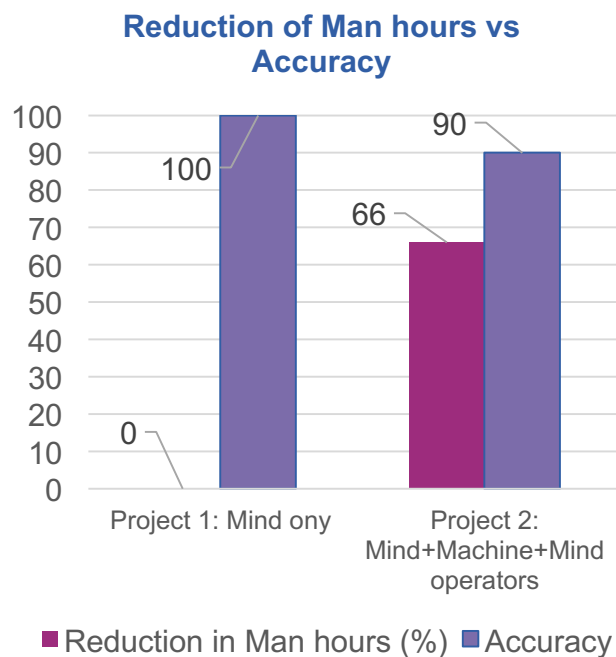
**"Machine+Mind operator" step -**

- For every classifier, identify and club/remove overlapping & not relevant technologies with help of assignees, classes and keyword based search string. This means modifying the input of KMX.

**"Machine" step -**

- Use KMX classifier to assign 'score' to each newly published patent to obtain automated categorization

**"Mind" step -**

- Manually categorize patents which were not categorized using KMX because they were falling under defined threshold range

### Reduction of Man hours vs Accuracy



Chart values:
- Project 1: Mind ony — 0 (Reduction in Man hours %), 100 (Accuracy)
- Project 2: Mind+Machine+Mind operators — 66 (Reduction in Man hours %), 90 (Accuracy)

■ Reduction in Man hours (%)  ■ Accuracy

## Benefit >

### Productivity

- **66% reduced effort of client**

### Quality

- **Achieved 90% accuracy**

# Case Study 4 – Competitive Benchmarking
## Implementing Mind+Machine for quick benchmarking study

**Organization**
Forest Industry Company

**Industry**
Pulp, Paper and Timber

**Function (s)**
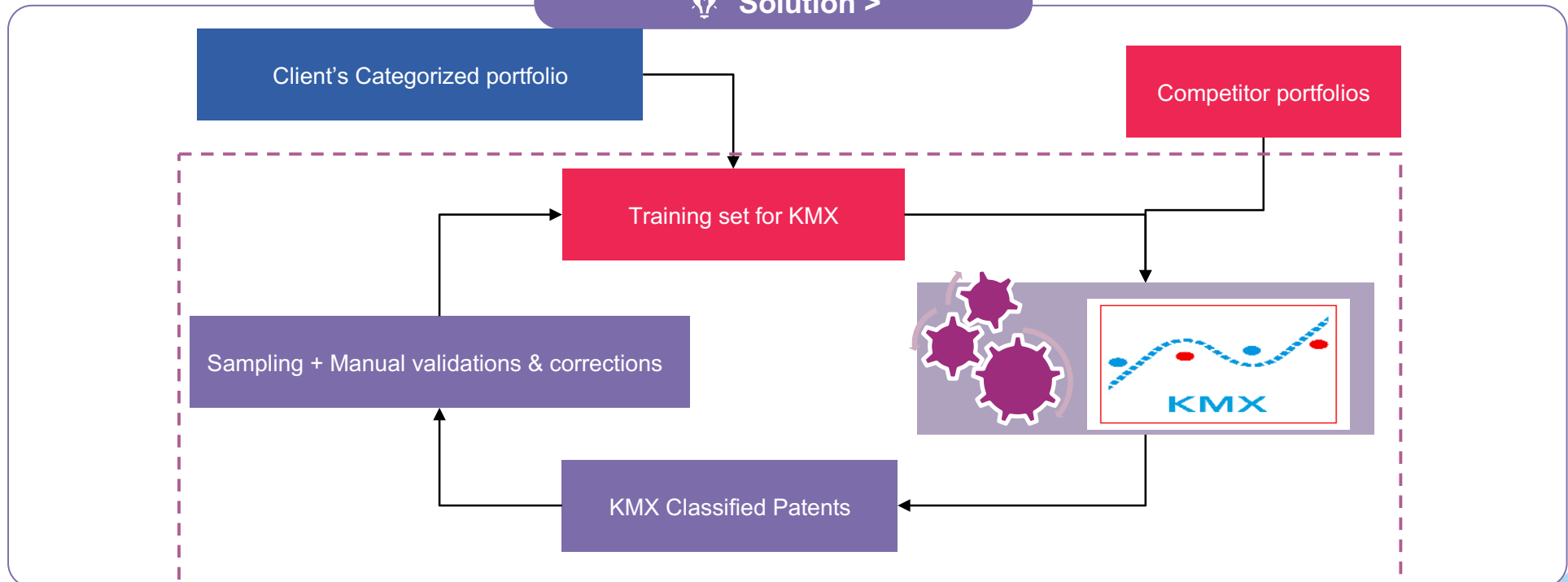R&D Strategy

**Geography**
HQ in Europe

**Business Challenge >**

- Client wanted to perform a quick competitive benchmark of its own portfolio with their competitor's, in select technical categories
- Client did not want 100% accuracy, but had limited budget for the study

**Solution >**

Client's Categorized portfolio

Competitor portfolios

Training set for KMX

Sampling + Manual validations & corrections

KMX

KMX Classified Patents

# Case Study 4 – Competitive Benchmarking
## Reduced 80% cost for client, while providing 70% accuracy
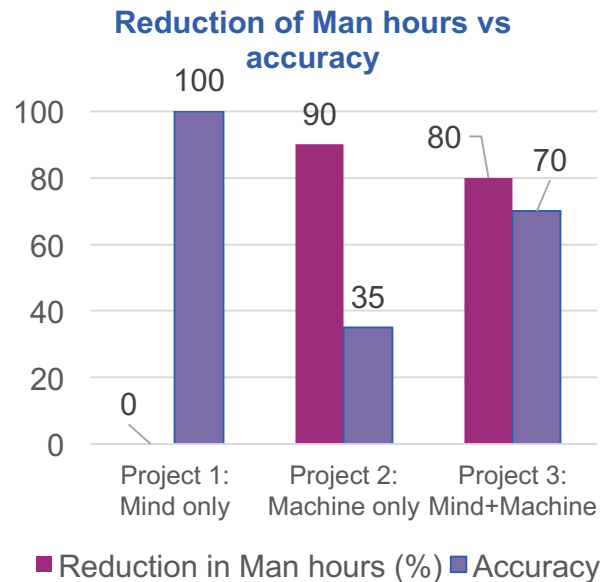
### Approach >

**Steps**:

**"Machine" steps -**

- Identify patents assigned to client and its competitors for last 5 years

- Use existing client's categorized patents to prepare KMX classifiers

- Categorization using KMX classifiers was done for ~40,000 extracted patents

**"Mind" step:**

- 4 iterations of manual checking of ~4,000 patents were performed to further tune KMX classifiers

**"Machine" step:**

- Modified KMX classifiers were run on entire set of 40K patents

**Reduction of Man hours vs accuracy**



Chart data:
- Project 1: Mind only — Reduction in Man hours (%): 0, Accuracy: 100
- Project 2: Machine only — Reduction in Man hours (%): 90, Accuracy: 35
- Project 3: Mind+Machine — Reduction in Man hours (%): 80, Accuracy: 70

■ Reduction in Man hours (%)  ■ Accuracy

### Benefit >

**Productivity**

- **80% reduced cost for client**

**Quality**

- **Ensured 70% accuracy – which was minimum expected from client**

# Case Study 5 – Extra review step in R&D alerts
## Implementing Mind+Machine for R&D Alerts

## Context >

**Organization**
Multinational Chemical Company

**Industry**
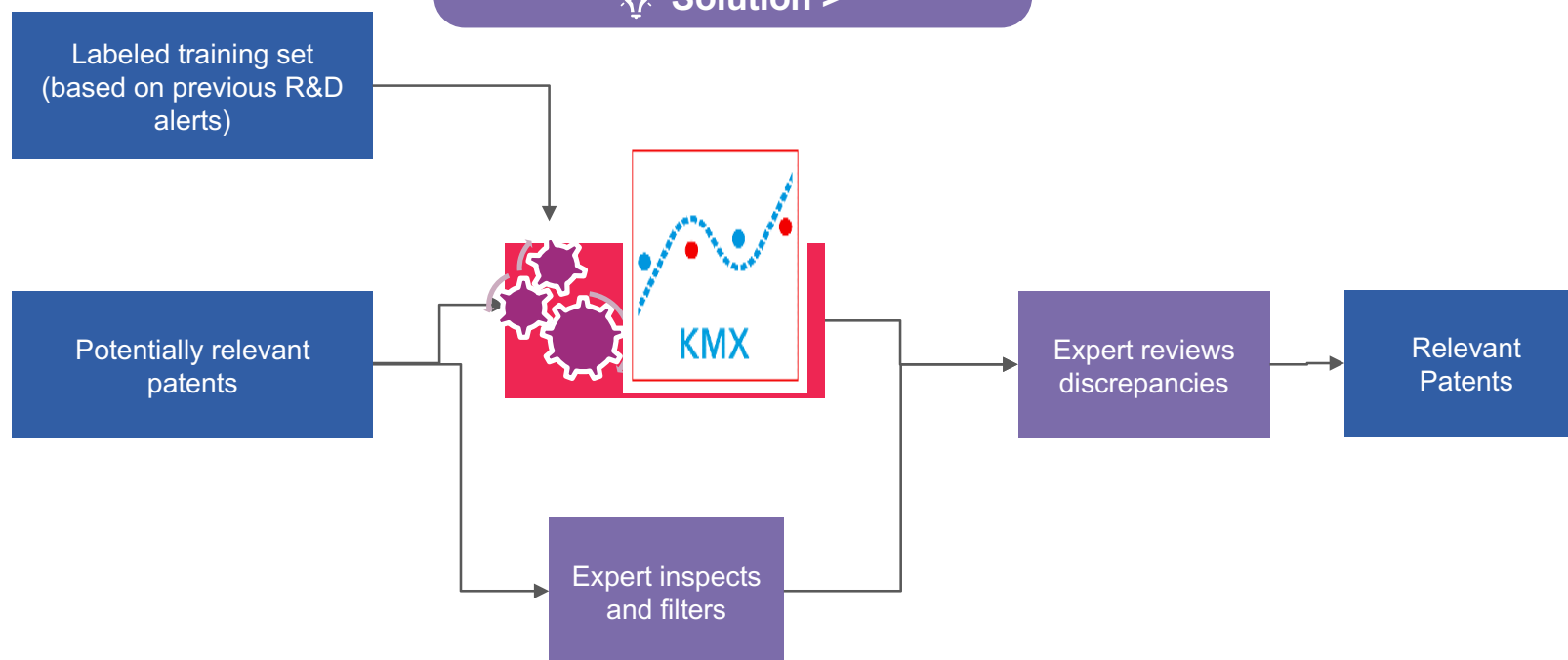Chemical and construction

**Function (s)**
R&D Alerts

**Geography**
HQ in Europe

## Business Challenge >

- Client maintained monthly R&D alerts, in which experts identified relevant new patents and applications in key technical domains.
- Client wanted to increase accuracy by adding an extra review step, without drastically increasing costs.

## Solution >

Labeled training set (based on previous R&D alerts)

Potentially relevant patents

KMX

Expert inspects and filters

Expert reviews discrepancies

Relevant Patents

# Case Study 5 – Extra review step in R&D alerts

Increased accuracy by 3% using Mind+Machine approach

## Approach >

**Steps:**

**"Machine" steps -**

One-time setup:

- Classifier is trained using previous R&D alert results
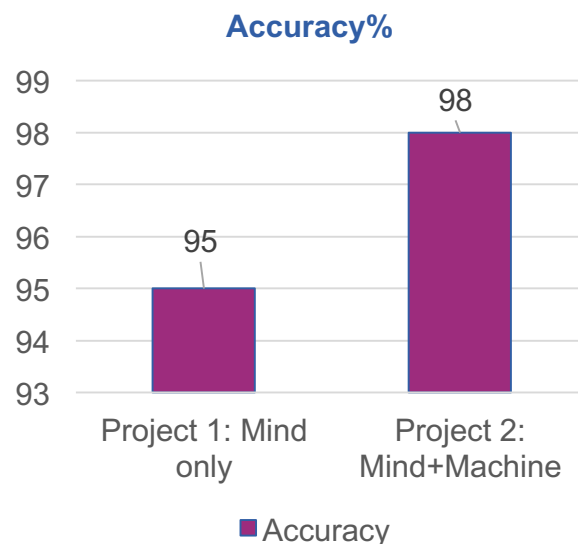
Monthly steps:

**"Mind" step:**

- Expert labels incoming patents

**"Machine" step:**

- KMX labels incoming patents

**"Mind" step:**

- Expert reviews patents for which no consensus was reached

**Accuracy%**

| | Project 1: Mind only | Project 2: Mind+Machine |
|---|---|---|
| Accuracy | 95 | 98 |

■ Accuracy

## Benefit >

**Productivity**

- **43% increased efficiency compared to workflow with an independent expert review**

**Quality**

- **Increased accuracy from 95% to 98%**

# Use case summary

| Case study | Mind+Machine | Accuracy | Efficiency gain |
|---|---|---|---|
| **Case Study 1 – Portfolio Categorization** | Machine only | 70% | 90% |
| **Case Study 2 – Proactive Patent Defence** | Machine only | 100%* | 35% |
| **Case Study 3 – Managing bi-annually portfolio updates** | Mind+Machine | 90% | 66% |
| **Case Study 4 – Competitive Benchmarking** | Mind+Machine | 70% | 80% |
| **Case Study 5 – Extra review step in R&D alerts** | Mind+Machine | 98% | 43% |

*Client obtained sufficient number of relevant patents