

FIAP

NBA

AGENDA

- AGILE / SCRUM
- CRISP-DM
- PERFIS DE PROFISSIONAL DE IA / FORMAÇÃO DE TIMES
- DEVOPS / AIOPS
- EVOLUÇÃO DE MODELOS DE ML EM PRODUÇÃO
- CANVAS COGNITIVO

+

+

•

•

•

•

•

□

•

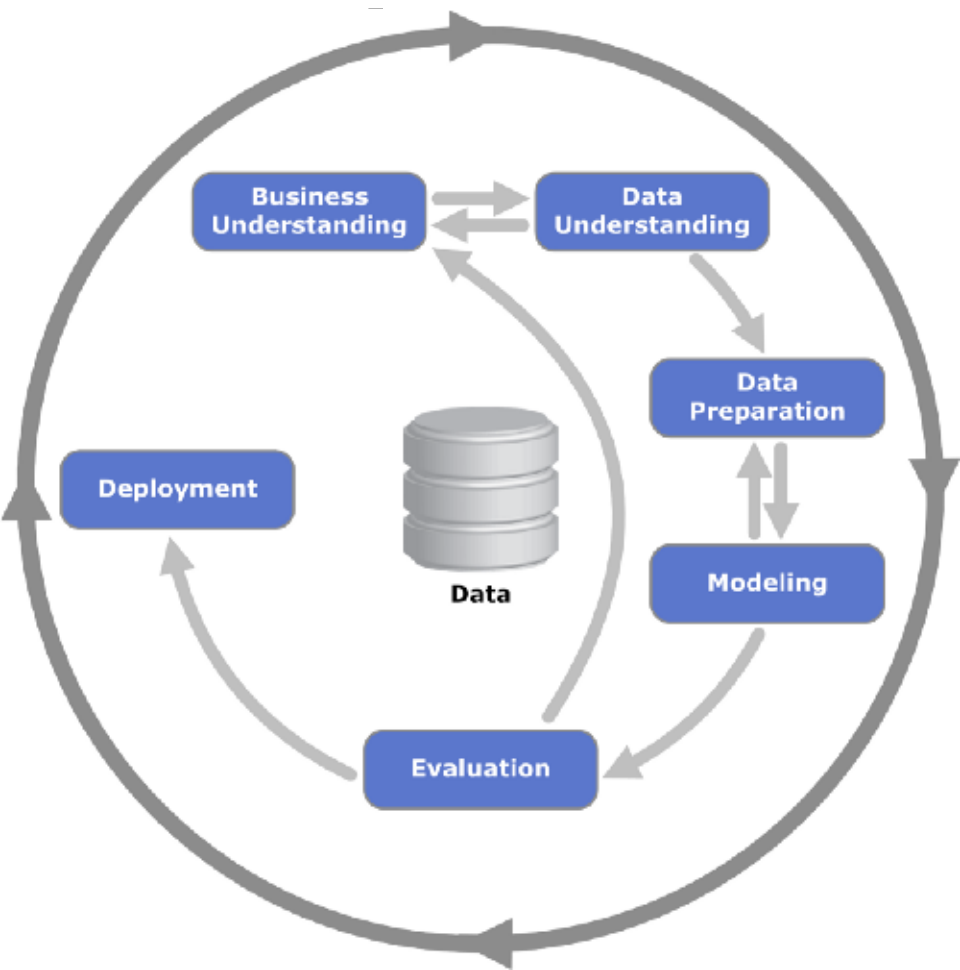
•

•

•

CRISP-DM





CRISP-DM

- O QUE É
- DE ONDE VEM
- COMO FUNCIONA
- ETAPAS
 - ENTENDIMENTO DO PROBLEMA
 - COMPREENSÃO DOS DADOS
 - PREPARAÇÃO DOS DADOS
 - MODELAGEM
 - AVALIAÇÃO
 - IMPLEMENTAÇÃO

O QUE É?

CRISP DM é a abreviação de Cross Industry Standard Process for Data Mining que, trazendo para o português, pode ser entendida como processo padrão da indústria cruzada para mineração de dados. Essa é uma metodologia capaz de transformar os dados da empresa em conhecimento e informações de gerenciamento.

PRÓS & CONTRAS

Prós: natureza flexível e cíclica; abordagem focada em tarefas; fácil de implementar.

Contras: não funciona para equipes; não cobre questões de comunicação.

ETAPAS

- ENTENDIMENTO DO PROBLEMA
- COMPREENSÃO DOS DADOS
- PREPARAÇÃO DOS DADOS
- MODELAGEM
- AVALIAÇÃO
- IMPLEMENTAÇÃO

ENTENDIMENTO DO PROBLEMA

A primeira coisa a ser feita é entender de fato qual o problema a ser resolvido, buscando todos os detalhes sobre o impacto dele na empresa e quais os objetivos em relação ao trabalho.

ENTENDIMENTO DO PROBLEMA

- **Background:** explique a situação da empresa e como o projeto vai ser direcionado para solucionar o problema;
- **Objetivo do projeto:** informe qual o objetivo maior que seu projeto tem;
- **Critério de sucesso:** deixe bem claro qual será a métrica que ditará se seu projeto atingiu o sucesso ou não.

COMPREENSÃO DOS DADOS

Essa etapa consiste em organizar e documentar todos os dados que se encontram disponíveis. É aqui que começa de fato o trabalho de mineração de dados, pois o profissional deve ser capaz de identificar quais são os dados importantes para a resolução do problema.

Nesse momento, o lado investigativo deve entrar em campo, para que os dados revelem problemas, soluções e tendências dos negócios.

PREPARAÇÃO DOS DADOS

Agora que os dados já foram identificados, documentados e analisados, é hora de aplicar a parte técnica de análise deles. Agora, serão preparadas as databases e definidos os formatos e questões técnicas da análise.

Nessa etapa, é feita a escolha dos dados que serão trabalhados e de como eles serão cruzados para resolver o problema da empresa.

PREPARAÇÃO DOS DADOS

- **Data Selection:** eliminação de outliers, colunas da tabela entre outros filtros, com as respectivas documentações de porque da inclusão/exclusão.
- **Data Cleaning:** Correção de formatos, datas, números como string, não padronização, etc.
- **Construct Data:** Adição de campos/colunas não disponíveis originalmente;
- **Integrating Data:** Junção de diferentes fontes de dados.

MODELAGEM

É nesta fase que são aplicadas de fato as técnicas de Data Mining, com base nos objetivos identificados no primeiro momento.

A partir de agora, a mineração de dados pode ser associada a análises preditivas, para que a empresa prepare-se para o futuro, resolvendo a questão principal.

Como? Os dados minerados podem ser usados para alimentar algoritmos que preveem as tendências dos negócios.

AVALIAÇÃO

Resultados vs Objetivos

Aplicação dos conhecimentos obtidos com o Data Mining.

Dados + Insights

IMPLEMENTAÇÃO

Entrada em produção

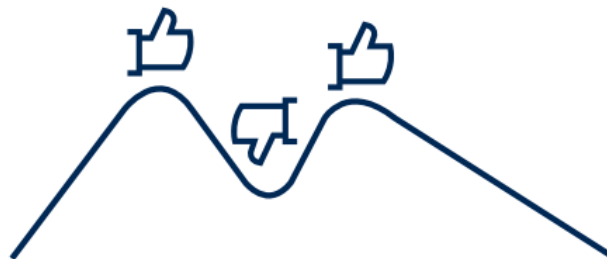
Mudanças de processos

Criação de produtos

Create an ROI Estimation

Detecting, avoiding, controlling bad events

- Churn
- Insurance claims
- Machine or process failures
- No-show of customers
- Complaints
- Default of a loan or payment
- Dangerous situations




Detecting or predicting good events


- Sales
- Renewals
- Shopping site locations
- Mineral resources

Costs of actions
(preventing a bad event)

Effectiveness of an action


**Costs incurred
by a bad event**

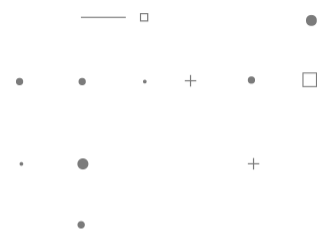



**Revenue generate
by a good event**

Long-term, softer and
compound effects

Adversarial effects of
a preventive or
facilitating action

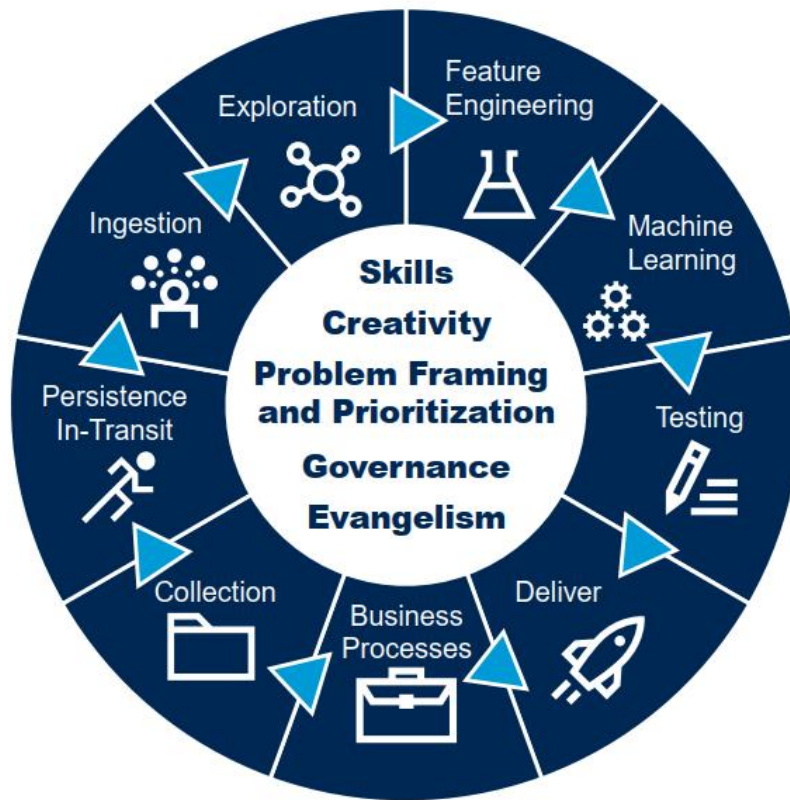
Cost of pattern recognition



PAPÉIS



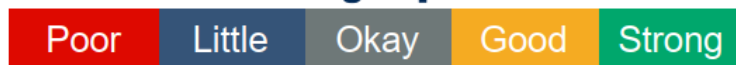
Data Science Is a Model Creation Cycle



Skill Map



Knowledge Spectrum



PAPÉIS (TDSP)

- Solution architect
- Project manager
- Data engineer
- Data scientist
- Application developer
- Project lead

Filter by title

Team Data Science Process

Documentation

Overview

Lifecycle

Lifecycle

1. Business understanding

2. Data acquisition and understanding

3. Modeling

4. Deployment

5. Customer acceptance

> Roles and tasks

Project planning

> Development

> Operationalization

> Worked out examples

Download PDF

The Team Data Science Process lifecycle

01/10/2020 • 2 minutes to read • +2

The Team Data Science Process (TDSP) provides a recommended lifecycle that you can use to structure your data-science projects. The lifecycle outlines the complete steps that successful projects follow. If you use another data-science lifecycle, such as the Cross Industry Standard Process for Data Mining (CRISP-DM), Knowledge Discovery in Databases (KDD), or your organization's own custom process, you can still use the task-based TDSP.

This lifecycle is designed for data-science projects that are intended to ship as part of intelligent applications. These applications deploy machine learning or artificial intelligence models for predictive analytics. Exploratory data-science projects and improvised analytics projects can also benefit from the use of this process. But for those projects, some of the steps described here might not be needed.

Five lifecycle stages

The TDSP lifecycle is composed of five major stages that are executed iteratively. These stages include:

1. Business understanding
2. Data acquisition and understanding

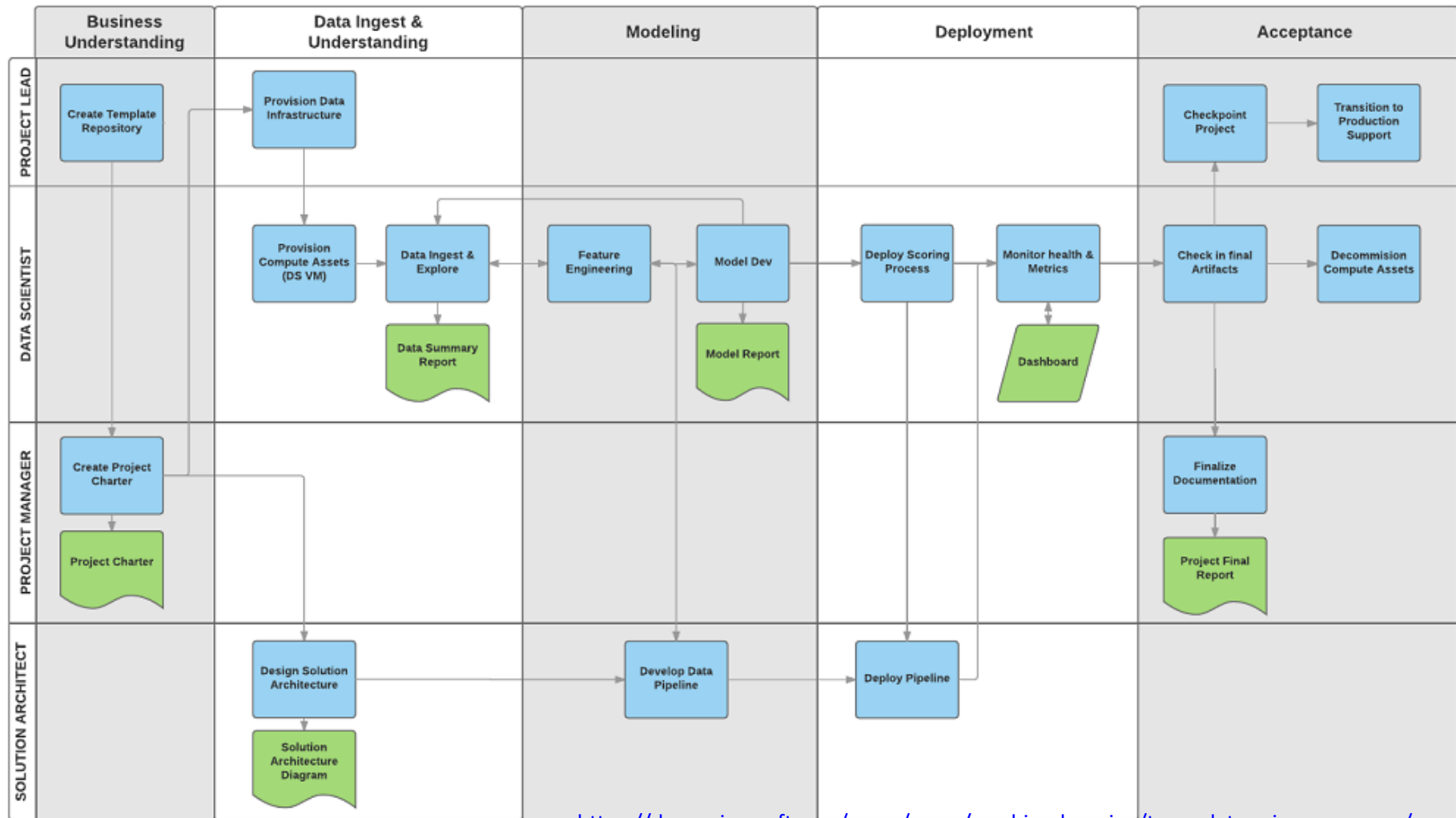
Is this page helpful?

Yes No

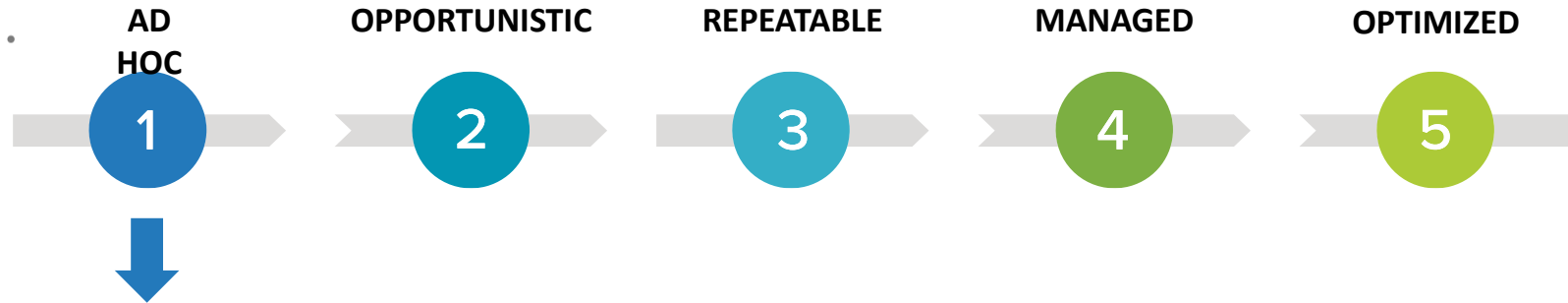
In this article

Five lifecycle stages

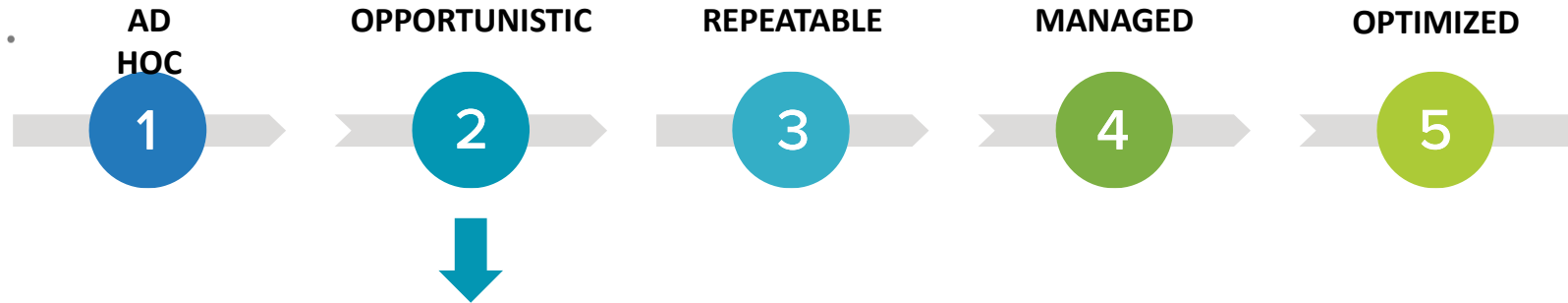
Next steps



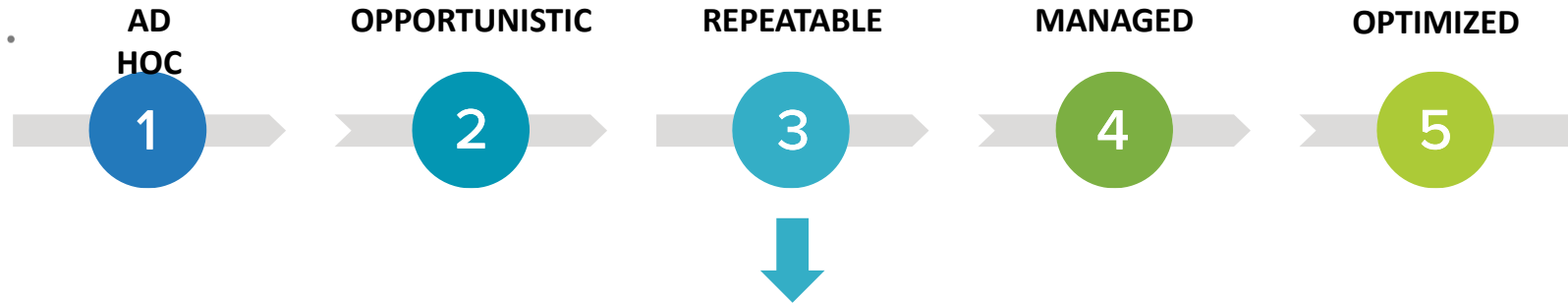
ESCALA DE MATURIDADE EM DX



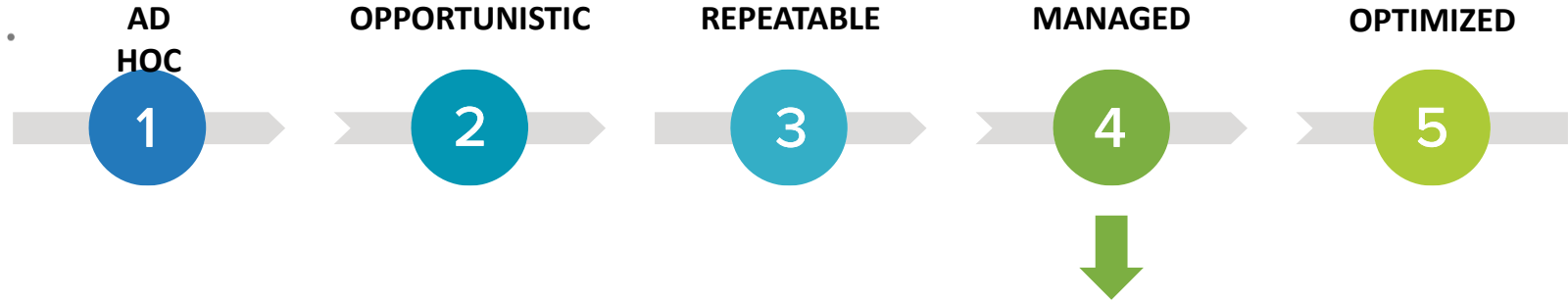
As metas de gestão para DX são mal definidas e eventualmente caóticas. O sucesso muitas vezes depende do esforço individual, e os benefícios não são amplamente compartilhados dentro do negócio. As iniciativas de negócios digitais e TI estão desconectadas e mal alinhadas com a estratégia corporativa e não focadas na experiência do cliente.



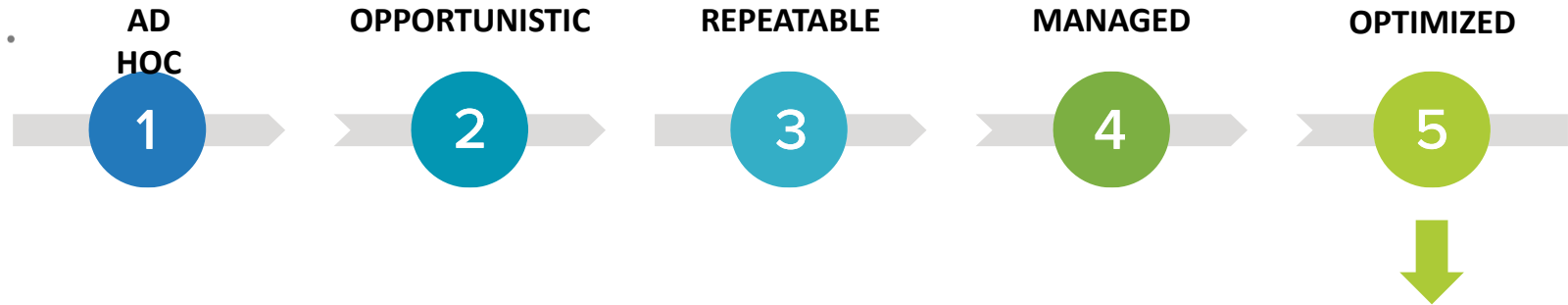
As capacidades básicas estão estabelecidas. As disciplinas necessárias para DX estão em vigor para repetir sucessos anteriores em iniciativas semelhantes. O negócio está para trás dos seus pares. A empresa identificou a necessidade de aprimorar digitalmente as estratégias de negócios, mas a execução é baseada em projetos isolados e o progresso não é previsível nem repetível.



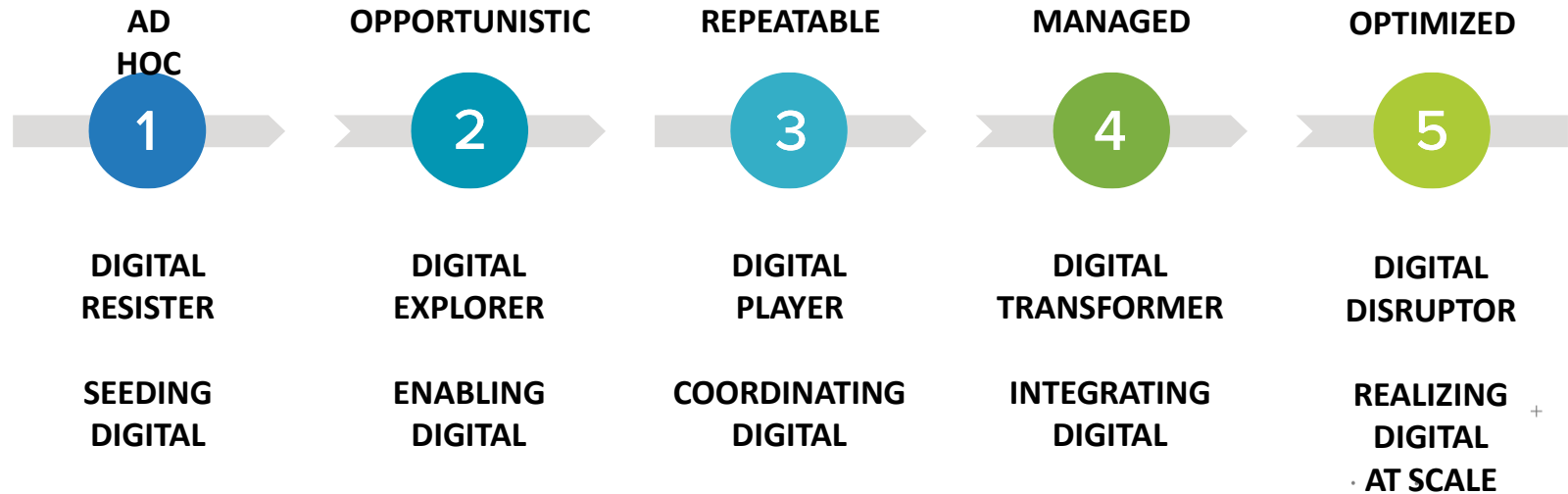
As metas de TI empresarial estão alinhadas no nível corporativo à estratégia de curto prazo e incluem iniciativas digitais de produtos e experiências de clientes, mas ainda não focadas no potencial disruptivo das iniciativas digitais. Os recursos são documentados, padronizados e integrados em nível corporativo. DX é um objetivo estratégico de negócios. O negócio está em pé de igualdade com os concorrentes e pares.



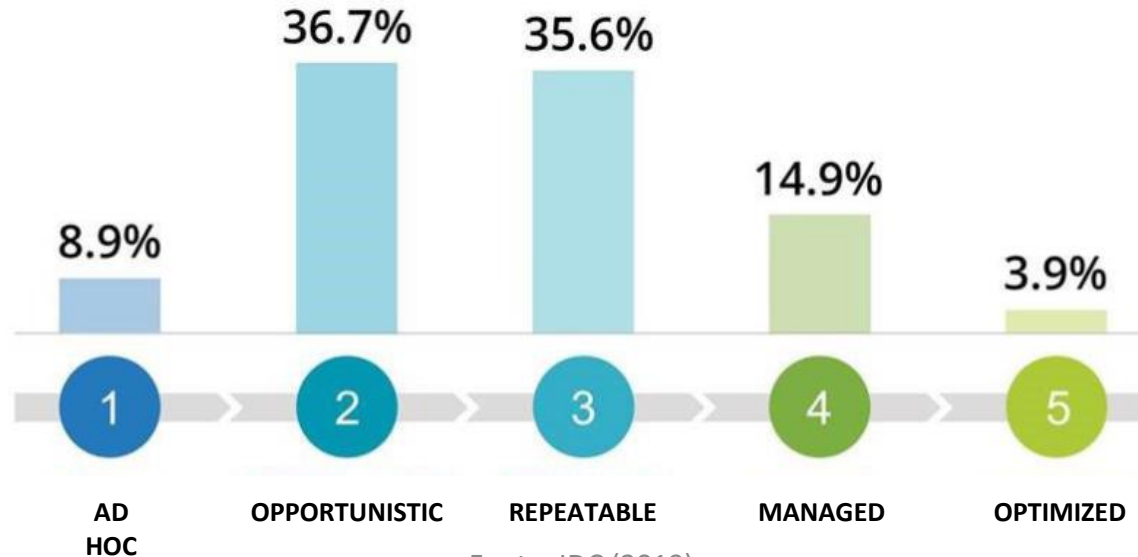
Os recursos para DX estão incorporados na empresa e fortemente ligados a uma visão ágil de gerenciamento. O negócio lidera seus pares e concorrentes. As disciplinas integradas e sinérgicas de gerenciamento de TI de negócios fornecem experiências de produto/serviço ativadas digitalmente de forma contínua.



A empresa é agressivamente disruptiva no uso de novas tecnologias digitais e modelos de negócios para impactar os mercados. A inovação empresarial é impulsionada pela alimentação constante do ecossistema e por feedback. A melhoria contínua é uma filosofia de gestão central. A liderança fomenta a tomada de riscos e a experimentação para desenvolver capacidades inovadoras.



The largest numbers of organizations are at stage 2 (Opportunistic). Such organizations are still operating on a project basis, resulting in inconsistent use of digitally enabled products and experiences and unpredictable progress. Becoming better aligned with business strategy and increased integration and consistency should be their priority. The next largest stage of maturity is stage 3 (Repeatable). While their digital products, services, and experiences are consistent, they do not truly innovate. **Innovation and integration should be the focus for organizations hoping to move from the repeatable to the managed stage.**



Fonte: IDC (2019)

OBRIGADO

 /vinicius-soares

FIAP

Copyright © 2018 | Professor (a) Vinicius Soares
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.

FIAP