



# FiTecserv



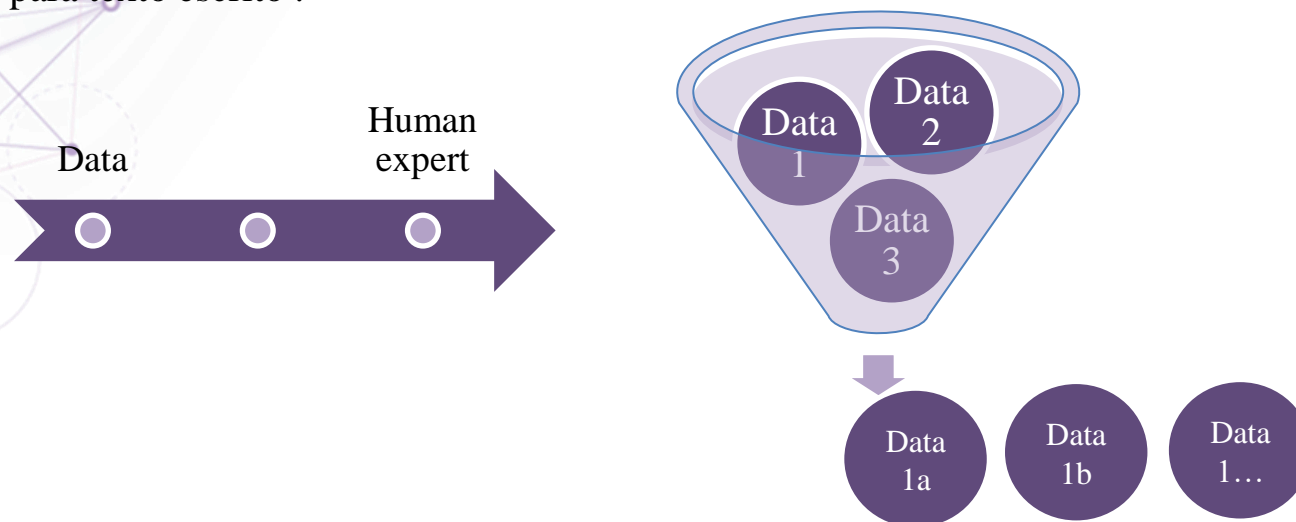
Data Augmentation

## Introdução

*Data Augmentation* refere-se às técnicas que são usadas para aumentar a disponibilidade de dados, fazendo alterações nos dados existentes e alterando as características dos mesmos. Esses dados são chamados de dados sintéticos recém-criados a partir de dados existentes e são de elevada importância no processo de treino de um modelo de aprendizagem máquina, pois disponibiliza dados em diversas formas que melhoram o processo de ciência de dados e o processo de aprendizagem máquina, especialmente em situações de escassez de dados .

Nossos serviços amplificam os dados disponíveis em diversas formas para vários modelos, sem necessidade de coletar novos dados. Essa técnica ajuda a economizar tempo e orçamento e é importante nos modelos de aprendizagem máquina, caso contrário, as organizações terão que gastar muito dinheiro na recolha de dados para iniciar modelos.

*Data Augmentation* é muito útil para *computer vision* (imagens) e situações de NLP, sendo muito benéfico para o som e também muito útil para texto escrito .



## Principais técnicas usadas

### Visão Computacional (Imagem)



### Visão Computacional:

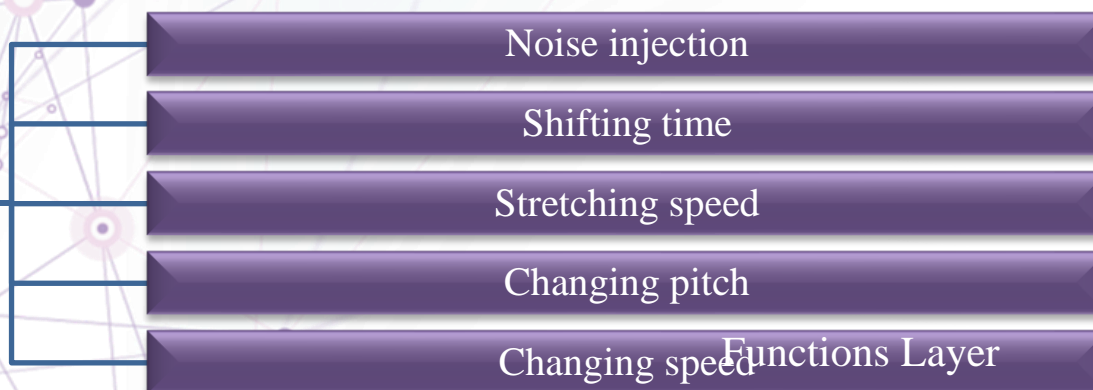
Nas soluções de visão computacional usamos entradas digitais como imagens, vídeos e outras entradas visuais para reunir informações significativas e, em seguida, usar essas informações derivadas para fazer recomendações úteis ou automatizar tarefas, aumentando assim a capacidade dos processos de negócio. Da mesma forma que a inteligência artificial permite que o sistema “pense”, a visão computacional permite que eles “vejam”, “observem” e “compreendam”.

Augmenting consiste em gerar novas imagens baseadas numa existente, modificando algumas características como: a cor, recortar ou ampliar.



## Principais técnicas usadas

NLP (Som)



*Augmenting* consiste em gerar novos sons baseados num já existente, modificando algumas características como: adicionar barulhos, mudar a velocidade ou a frequência.

**NLP (Natural Language Processing) ou Processamento Natural de Linguagem:**

Com esta tecnologia facilitamos a comunicação entre computadores e entre computador/humano usando linguagem humana. Basicamente, é uma maneira pela qual os computadores são programados para analisar e processar uma enorme quantidade de dados linguísticos. Um exemplo do uso dessa tecnologia é o OCR (Optical Character Recognition), em que se economiza muito tempo no processamento de informações digitais, obtendo resultados significativos, usando depois essas informações derivadas para ajudar a tomar a decisão certa no momento certo. Outros exemplos são a automatização de transcrição e tradução de texto ou a utilização de voz sintética em centros de interação (e.g. Centros de contato com cliente)

## Principais técnicas usadas

NLP (Texto)

Synonym replacement

Word insertion

Word swap

Word deletion

Back translation

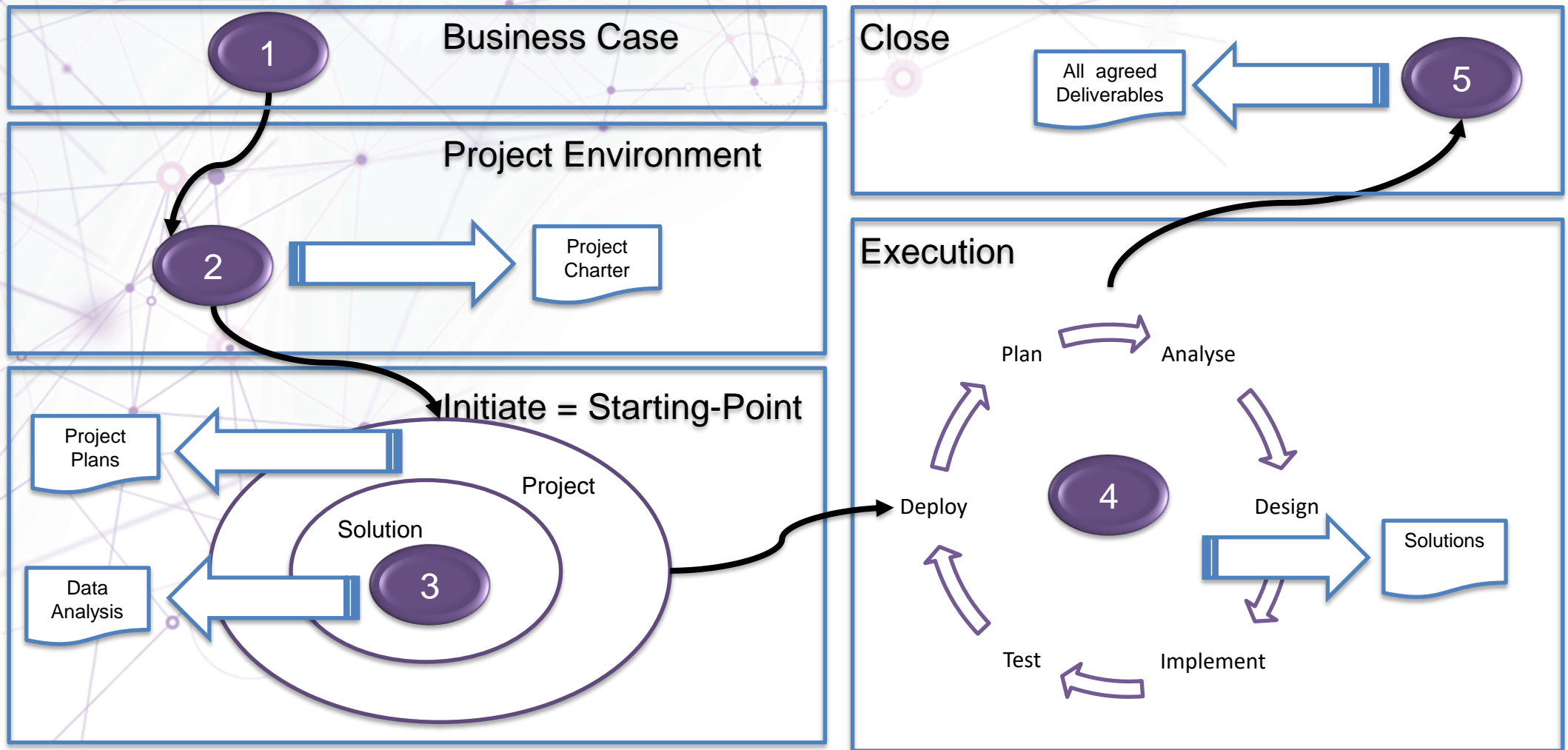
Saw contextualized word embeddings

*Augmenting* consiste em gerar novos textos baseados em textos pré-existentes, modificando algumas características como: adicionar ou remover palavras, alterar palavras pelos seus sinónimos.

**NLP (Natural Language Processing) ou Processamento Natural de Linguagem:**

Com esta tecnologia facilitamos a comunicação entre computadores e entre computador/humano usando linguagem humana. Basicamente, é uma maneira pela qual os computadores são programados para analisar e processar uma enorme quantidade de dados linguísticos. Um exemplo do uso dessa tecnologia é o OCR (Optical Character Recognition), em que se economiza muito tempo no processamento de informações digitais, obtendo resultados significativos, usando depois essas informações derivadas para ajudar a tomar a decisão certa no momento certo. Outros exemplos são a automatização de transcrição e tradução de texto ou a utilização de voz sintética em centros de interação (e.g. Centros de contato com cliente)

## Modelo Implementação AitecServ (AIM)



## AIM – Caso de Negócio & Ambiente Projeto

### Caso de Negócio

Identificar entidades envolvidas  
(Key: Identificar Donos dos dados)

Preparação/Identificação Caso de negócio  
(Que dados se pretendem tratar?)

Visão Projeto  
(Quais os Benefícios esperados?)

Requisitos de Alto Nível  
(Identificar as técnicas mais importantes)

### Ambiente Projeto

Identificar requisitos de conformidade e privacidade (Imagens com pessoas e/ou marcas?)

Identificar requisitos de acesso aos dados

Endereçar acessos e autorizações

Capturar pressupostos, constrangimentos e acordos prévios



## AIM – Starting Point

### **Gestão:**

Neste fase a principal preocupação é de elaborar os diferentes planos do projeto. Estes projetos requerem normalmente planeamento claro e detalhado de forma a garantir o sucesso do mesmo.

Este tipo de projeto caracterize-se por equipas muito pequenas e de baixa complexidade nas componentes de gestão.

### **Solução:**

Nesta fase avaliamos a disponibilidade dos dados e executamos uma primeira análise exploratória e bastante descritiva, de forma a inferir questões de qualidade dos dados e capturar alguns padrões e conhecimento que nos guiem na próxima fase.

## Gestão

*Identificar principais entregáveis*

*Identificar Riscos*

*Preparação inicial dos Planos do Projeto*

## Solução

*Avaliar disponibilidade de dados*

*Análise de dados Exploratória*



## AIM – Execução

### Gestão

*Monitorizar e  
Controlar os  
objetivos Projeto*

*Gerir os diversos  
planos projeto*

*Gerir o Backlog de  
Produto e dos  
Sprints*

### Solução

*Implementação  
segundo princípios  
AGILE*

*Primeira entrega de  
piloto com “Go-No  
Go”*

*Incremental e  
Iterativo*

#### **Gestão:**

A atividade principal será de monitorizar e controlar o desempenho e os objetivos do projeto, efetuando os ajustamentos necessários de forma a manter o projeto em linha com o planeado.

#### **Solução:**

Os nossos serviços seguem preferencialmente uma abordagem AGILE, entregando numa primeira iteração um conjunto de dados gerados para efeitos de avaliação de qualidade e endereçar os objetivos pretendidos e avaliar as técnicas e ferramentas usadas..



# Atecserv

Obrigado

Para mais informação ou para marcar  
uma reunião contacte-nos em:



<http://www.atecserv.pt/contactos>  
<mailto:geral@atecserv.pt>