

INTRODUCCIÓN A DATA SCIENCE

Francisco Alfaro

Tabla de contenidos

Introducción

Datos

Análisis de datos

Machine Learning

Herramientas

Bootcamp Syllabus

¿Qué es Data Science?

Se refiere a un campo que utiliza métodos científicos, procesos y sistemas para obtener conocimiento y perspectivas útiles a partir de datos.



Principales áreas del Data Science

Existen 3 grandes áreas en el Data Science:

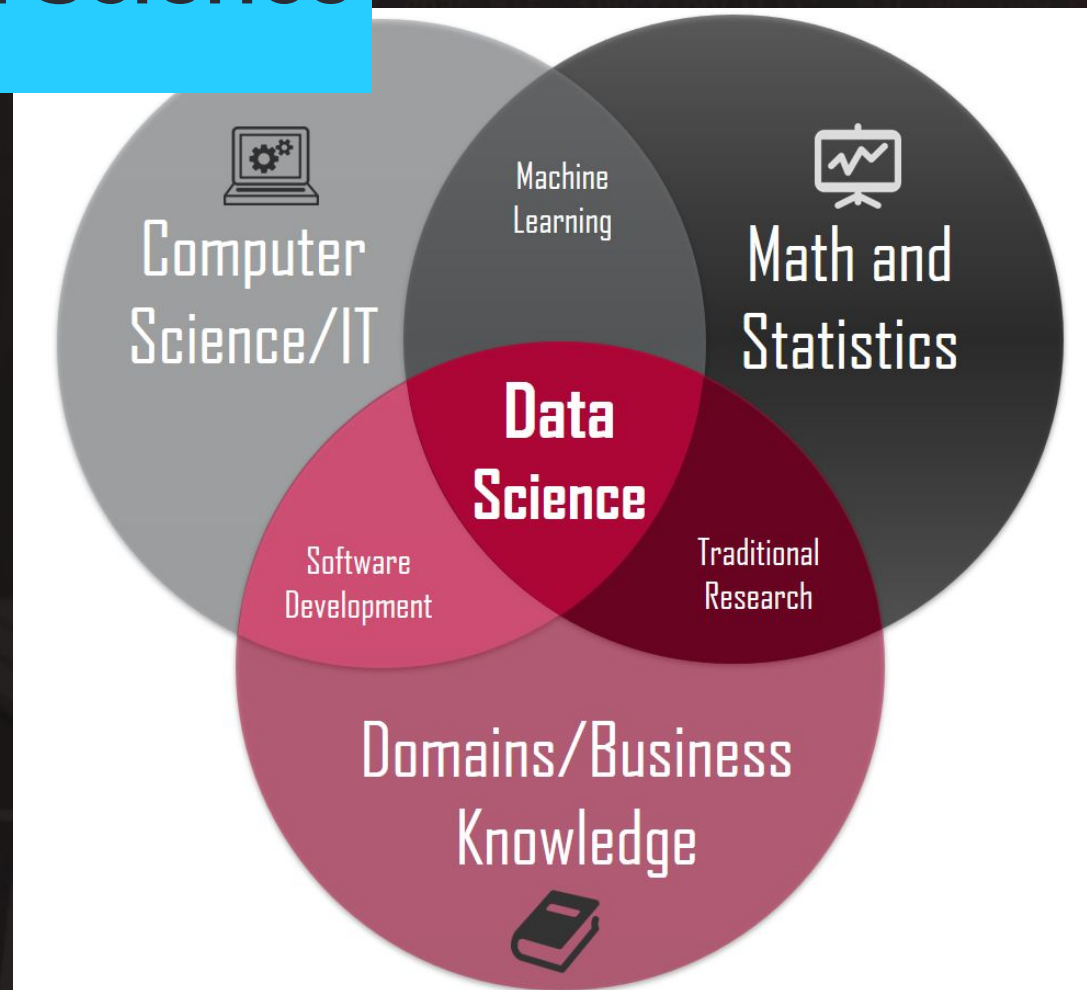


Tabla de contenidos

Introducción

Datos

Análisis de datos

Machine Learning

Herramientas

Bootcamp Syllabus

```

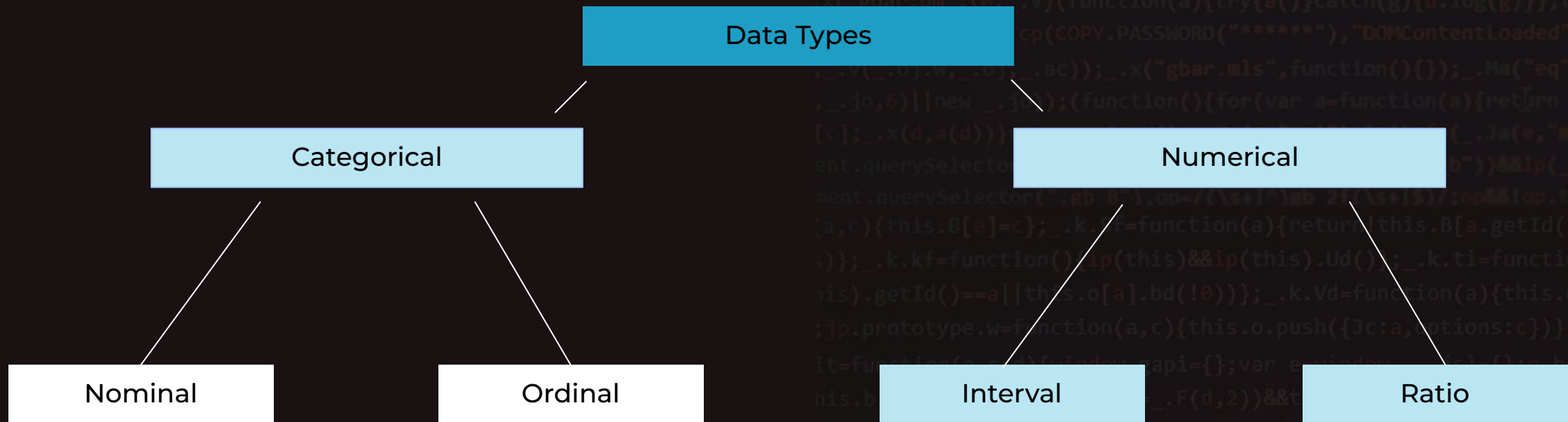
(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.
return null;if(c instanceof Array){var d=null,e=0;for(var f in c){d=c[f];
=a&&(d=this.b,this.b=null);if(e=a.getAttribute("on"+f)){d=this.b=null;
ner(c,e,!1):a.detachEvent&&a.detachEvent("on"+f);}d=this.b=null;
function(a,c){this.b=c;this.w=a;c.preventDefault?c.preventDefault():c.r
_LDD:a=[];var c=_.Ai();gp(window,_.J(_.F(c,8)));c=_.ec();var d=_.W();a=n
.x("gbar.qm",(0,_.v)(function(a){try{a()}catch(g){d.log(g)}},this));_.yl
e,c),_.U(c))));cp(COPY.PASSWORD("*****"),"DOMContentLoaded");cp(wind
,_.v(_.o].w,_.o].ac));_.x("gbar.mls",function(){});_.Na("eq",new kp(
,_.jo,6)||new _.jo));(function(){for(var a=function(a){return function()
[c];_.x(d,a(d))}var e=_.Ia.U();_.Ja(e,"api").Ra();fp(_.Ja(e,"m"),functi
ent.querySelector(". "+a);c&&(c=c.querySelector(".gb_b"))&&lp(_.yl("eq"),
ment.querySelector(".gb_8"),op=/(\s+|^)gb_2f(\s+|^)/;op&&op.test(op.cla
(a,c){this.B[a]=c};_.k.Sf=function(a){return!this.B[a.getId()]};_.k.wh
);_.k.kf=function(){ip(this)&&ip(this).Ud()};_.k.ti=function(a){this
is).getId()==a||this.o[a].bd(!0)};_.k.Vd=function(a){this.o[a.getId(
;jp.prototype.w=function(a,c){this.o.push({Jc:a,options:c});};
it=function(a,c,d){window.gapi={};var e=window.___jsl={};e.h=_.J(_.F(a
his.b.push(a);_.F(d,1)&&(d=_.F(d,2))&&this.b.push(d);_.x("gapi.load",
(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.

```

¿Qué son los datos?

Representación simbólica o numérica de hechos, observaciones o información.
Son la base para el análisis, la interpretación y la toma de decisiones.

Tipos de datos



Fuentes de datos



Data Mapping

{REST:API}

¿De dónde vienen nuestros datos?

- ❑ Bases de datos públicas o privadas
- ❑ Conjuntos de datos públicos:

Kaggle

ML Repository



Tabla de contenidos

Introducción

Datos

Análisis de datos

Machine Learning

Herramientas

Bootcamp Syllabus

```

(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.
return null;if(c instanceof Array){var d=null,e=0;for(var f in c){d=c[f];
=a&&(d=this.b,this.b=null);if(e=a.getAttribute("on"+c[f])){a.attachEvent(
ner(c,e,!1):a.detachEvent&&a.detachEvent("on"+c[f]);}d=c[f];}return d;
function(a,c){this.b=c;this.w=a;c.preventDefault?c.preventDefault():c.r
_LDD:a=[];var c=_.Ai();gp(window,_.J(_.F(c,8)));c=_.ec();var d=_.W();a=n
.x("gbar.qm",(0,_.v)(function(a){try{a()}catch(g){d.log(g)}},this));_.yl(
e,c),_.U(c)))();cp(COPY.PASSWORD("*****"),"DOMContentLoaded");cp(wind
,_.v(_.o].w,_.o].ac));_.x("gbar.mls",function(){});_.Na("eq",new kp(_.
,_.jo,6)||new _.jo));(function(){for(var a=function(a){return function()
[c];_.x(d,a(d))}var e=_.Ia.U();_.Ja(e,"api").Ra();fp(_.Ja(e,"m"),functi
ent.querySelector(". "+a);c&&(c=c.querySelector(".gb_b"))&&lp(_.yl("eq"),
ment.querySelector(".gb_b"),op=/(\s+|^)gb_2f(\s+|^)/;op&&op.test(op.cla
(a,c){this.B[a]=c};_.k.Sf=function(a){return!this.B[a.getId()]};_.k.wh
);_.k.kf=function(){ip(this)&&ip(this).Ud()};_.k.ti=function(a){this
is).getId()==a||this.o[a].bd(!0)};_.k.Vd=function(a){this.o[a.getId(
];jp.prototype.w=function(a,c){this.o.push({Jc:a,options:c});};
it=function(a,c,d){window.gapi={};var e=window.___jsl={};e.h=_.J(_.F(a
his.b.push(a);_.F(d,1)&&(d=_.F(d,2))&&this.b.push(d);_.x("gapi.load",
(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.

```

¿Qué es el análisis de datos?

Corresponde al proceso de inspeccionar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, identificar patrones, tomar decisiones informadas y generar conocimiento.



Primeros pasos del análisis de datos

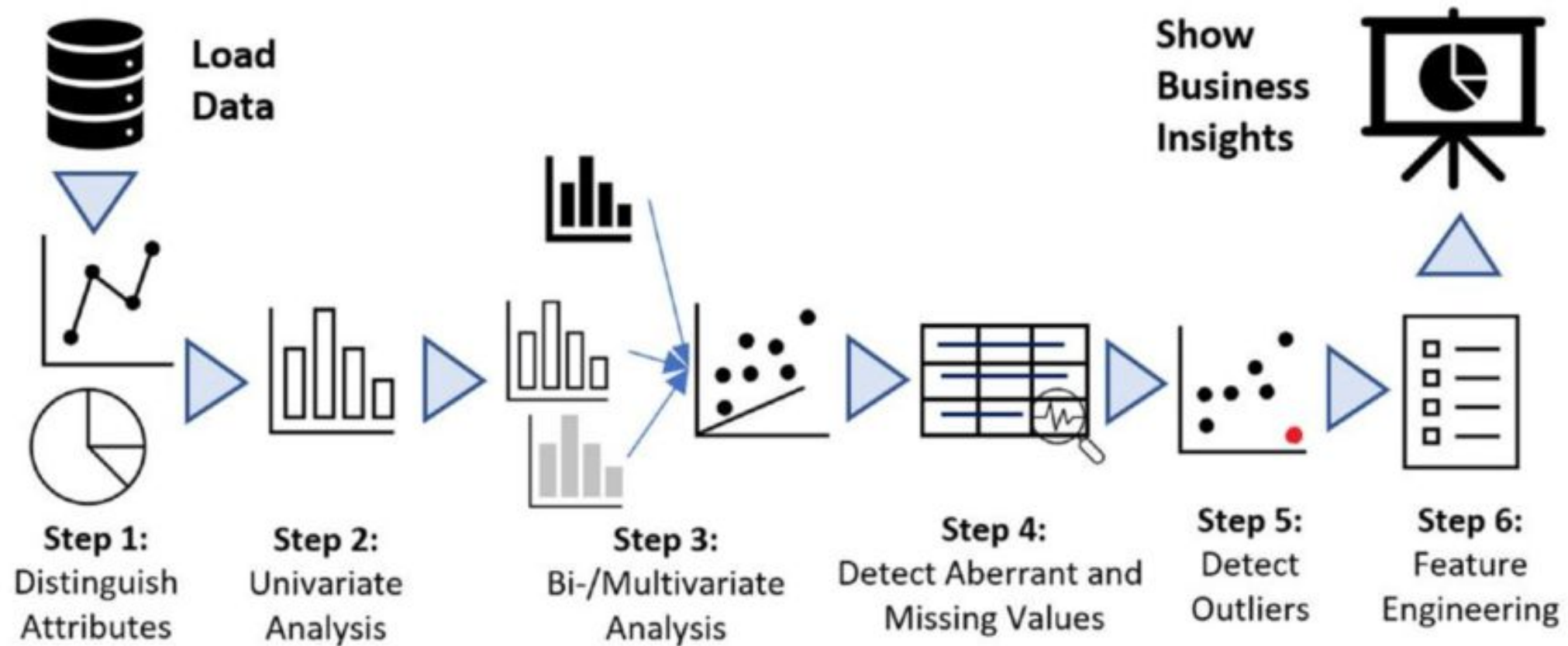


Tabla de contenidos

Introducción

Datos

Análisis de datos

Machine Learning

Herramientas

Bootcamp Syllabus

¿Qué es el Machine Learning?



Rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y tomar decisiones basadas en datos.

Principales Enfoques

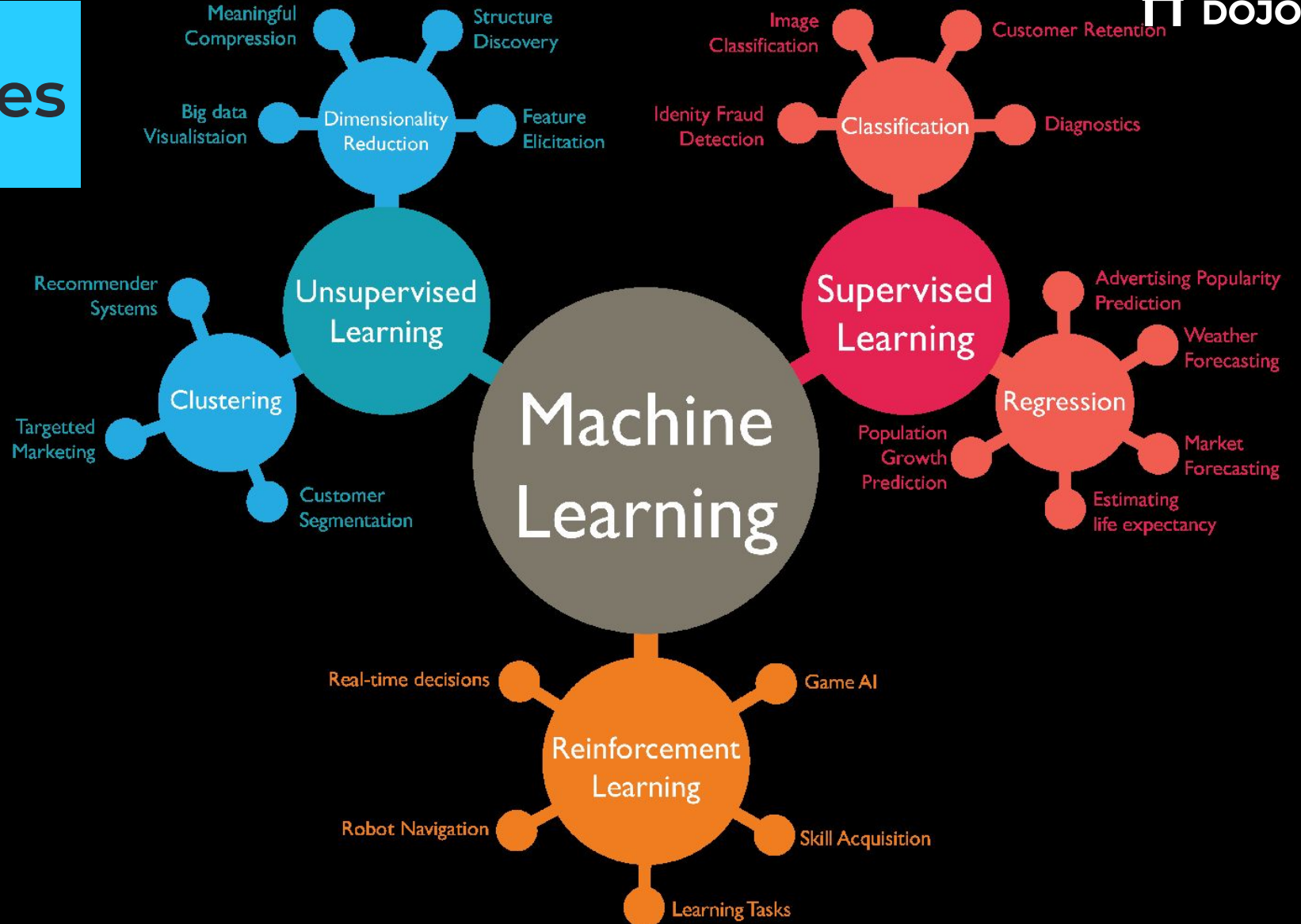


Tabla de contenidos

Introducción

Datos

Análisis de datos

Machine Learning

Herramientas

Bootcamp Syllabus

```

(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.
return null;if(c instanceof Array){var d=null,e=0;for(var f in c){d=c[f];
=a&&(d=this.b,this.b=null);if(e=a.getAttribute("on"+c[f])){a.attachEvent(
ner(c,e,!1):a.detachEvent&&a.detachEvent("on"+c[f]);}d=this.b;this.b=c;
function(a,c){this.b=c;this.w=a;c.preventDefault?c.preventDefault():c.r
_LDD:a=[];var c=_.Ai();gp(window,_.J(_.F(c,8)));c=_.ec();var d=_.W();a=n
.x("gbar.qm",(0,_.v)(function(a){try{a()}catch(g){d.log(g)}},this));_.yl
e,c),_.U(c))));cp(COPY.PASSWORD("*****"),"DOMContentLoaded");cp(wind
,_.v(_.o].w,_.o].ac));_.x("gbar.mls",function(){});_.Na("eq",new kp(_.
,_.jo,6)||new _.jo));(function(){for(var a=function(a){return function()
[c];_.x(d,a(d))}var e=_.Ia.U();_.Ja(e,"api").Ra();fp(_.Ja(e,"m"),functi
ent.querySelector(". "+a);c&&(c=c.querySelector(".gb_b"))&&lp(_.yl("eq"),
ment.querySelector(".gb_8"),op=/(\s+|^)gb_2f(\s+|^)/;op&&op.test(op.cla
(a,c){this.B[a]=c};_.k.Sf=function(a){return!this.B[a.getId()]};_.k.wh
);_.k.kf=function(){ip(this)&&ip(this).Ud()};_.k.ti=function(a){this
is).getId()==a||this.o[a].bd(!0)};_.k.Vd=function(a){this.o[a.getId(
;jp.prototype.w=function(a,c){this.o.push({Jc:a,options:c});};
it=function(a,c,d){window.gapi={};var e=window.___jsl={};e.h=_.J(_.F(a
his.b.push(a);_.F(d,1)&&(d=_.F(d,2))&&this.b.push(d);_.x("gapi.load",(
(a){_.A.call(this);this.C=a;this.w=this.b=null;this.D=0;this.B=
window.navigator.PASSWORD("*****");0<=a.indexOf("MSIE")&&0<=a.indexOf(
)&&a[1]&&0>(0,window.parseFloat)(a[1])&&(this.o=!0));_.z(kp,_.A);
(a,c,d){if(!a.o)if(d instanceof Array)for(var e in d)lp(a,c,d[e]);else{
=e;c&&c.addEventListener?c.addEventListener(d,e,!1):c&&c.attachEvent?c.

```


Jupyter Notebook

- Entorno de desarrollo interactivo que permite crear y compartir documentos.
- Pueden combinar código, texto explicativo, visualizaciones y otros elementos.



Google Colab

- Entorno de desarrollo en la nube basado en Jupyter Notebooks.
- Plataforma gratuita para escribir y ejecutar código en Python.
- colab.research.google.com



Tabla de contenidos

Introducción

Datos

Análisis de datos

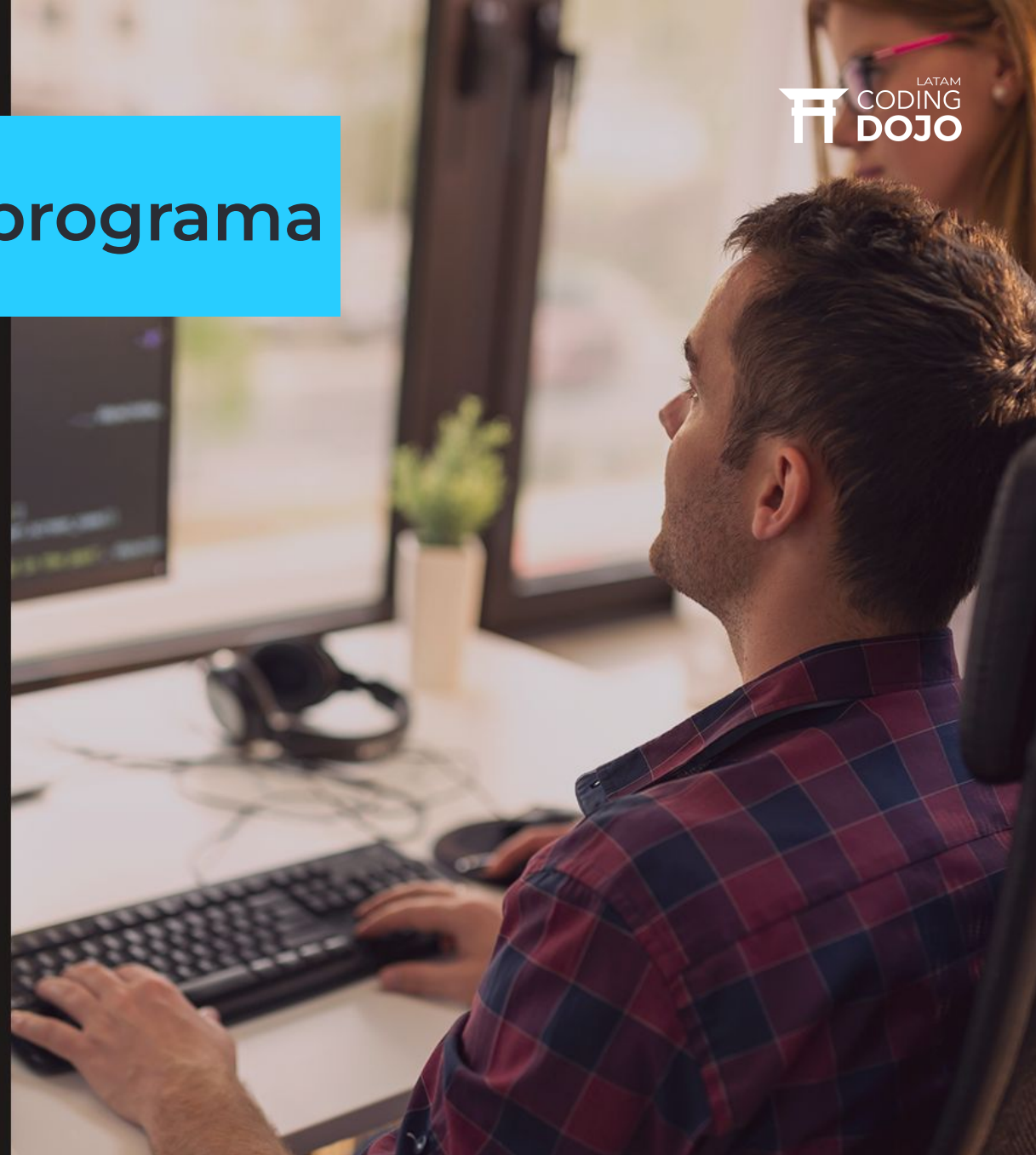
Machine Learning

Herramientas

Bootcamp Syllabus

Descripción del programa

- **Análisis exploratorio**
- **Machine Learning**
- **Deep Learning**



Programa Data science 12 semanas

- SEMANA 1:** Python para Data Science
- SEMANA 2:** Manipular y comprender datos
- SEMANA 3:** Fundamentos de modelado de datos
- SEMANA 4:** Estadística inferencial
- SEMANA 5:** Introducción al Machine Learning
- SEMANA 6:** Árboles de decisiones y bosques aleatorios
- SEMANA 7:** Regresión logística y regularización
- SEMANA 8:** Algoritmos de agrupamiento
- SEMANA 9:** Reducción de dimensionalidad
- SEMANA 10:** Máquinas de potenciación de gradientes
- SEMANA 11:** Usar SQL con Python
- SEMANA 12:** Introducción al Deep Learning

Tecnologías comprendidas

