

# Python数据分析与挖掘实战

## 1、快速了解数据分析与挖掘技术

韦玮

# 自我介绍

- 天善商业智能和大数据社区 Python 讲师 – 韦玮
- 天善社区 ID - 韦玮
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区 Python 版块。

# 本课概要

- 什么是数据分析与挖掘技术
- 数据分析与挖掘技术能做什么事情？
- 数据挖掘的过程

# 什么是数据分析与挖掘技术

所谓数据分析，即对已知的数据进行分析，然后提取出一些有价值的信息，比如统计出平均数、标准差等信息，数据分析的数据量有时可能不会太大，而数据挖掘，是指对大量的数据进行分析与挖掘，得到一些未知的，有价值的信息等，比如从网站的用户或用户行为数据中挖掘出用户的潜在需求信息，从而对网站进行改善等。

数据分析与数据挖掘密不可分，数据挖掘是数据分析的提升。

# 数据分析与挖掘技术能做什么事情？

数据挖掘技术可以帮助我们更好的发现事物之间的规律。所以，我们可以利用数据挖掘技术实现数据规律的探索，比如发现窃电用户、发掘用户潜在需求、实现信息的个性化推送、发现疾病与症状甚至疾病与药物之间的规律.....等。

# 数据挖掘的过程

数据挖掘的过程主要有：

- 1、定义目标
- 2、获取数据（常用的手段有通过爬虫采集或者下载一些统计网站发布的数据）
- 3、数据探索
- 4、数据预处理（数据清洗【去掉脏数据】、数据集成【集中】、数据变换【规范化】、数据规约【精简】）
- 5、挖掘建模（分类、聚类、关联、预测）
- 6、模型评价与发布

# Python数据分析与挖掘实战

## 2、数据分析与挖掘相关模块简介 与安装 韦玮

# 自我介绍

- 天善商业智能和大数据社区 Python 讲师 – 韦玮
- 天善社区 ID - 韦玮
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区 Python 版块。



# 本课概要

- 相关模块简介
- 相关模块的安装与技巧
- 相关模块的基本使用

# 相关模块简介

- 1、numpy 可以高效处理数据、提供数组支持、很多模块都依赖他，比如pandas、scipy、matplotlib都依赖他，所以这个模块是基础。
- 2、pandas 我们课程后续用得最多的一个模块，主要用于进行数据探索和数据分析。
- 3、matplotlib 作图模块，解决可视化问题。
- 4、scipy 主要进行数值计算，同时支持矩阵运算，并提供了很多高等数据处理功能，比如积分、傅里叶变换、微分方程求解等。
- 5、statsmodels 这个模块主要用于统计分析
- 6、Gensim 这个模块主要用于文本挖掘
- 7、sklearn、keras 前者机器学习，后者深度学习

# 相关模块的安装与技巧

模块安装的顺序与方式建议如下：

- 1、numpy、mkl（下载安装）
- 2、pandas（网络安装）
- 3、matplotlib（网络安装）
- 4、scipy（下载安装）
- 5、statsmodels（网络安装）
- 6、Gensim（网络安装）

# 相关模块的基本使用

接下来为大家通过实战讲解相关模块的基本使用。

# Python数据分析与挖掘实战

## 3、Python数据导入实战

韦玮

# 自我介绍

- 天善商业智能和大数据社区 Python 讲师 – 韦玮
- 天善社区 ID - 韦玮
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区 Python 版块。

# 本课概要

- 导入csv数据
- 导入excel数据
- 导入MySQL数据库里的数据
- 导入html数据
- 导入文本数据

# 导入csv数据

csv是一种常见的数据存储格式，基本上我们遇到的数据都可以转为这种存储格式。在Python数据分析中，我们可以使用pandas模块导入csv数据，接下来通过实战讲解。



# 导入excel数据

excel是一种表格文件，在Python数据分析中，我们同样可以使用pandas模块导入excel表格里面的数据，接下来通过实战讲解。

# 导入MySQL数据库里的数据

MySQL是一种很常见的数据库，在Python数据分析中，我们也可以直接从MySQL中取数据，接下来通过实战讲解。

# 导入html数据

使用pandas，可以直接从html网页中加载对应table表格中的数据，但是在使用read\_html()之前，需要先安装html5lib模块与beautifulsoup4模块。接下来通过实战讲解。

# 导入文本数据

有时我们希望直接导入txt文本数据，其实也是可以的，接下来通过实战讲解。