

利用数据分析 识别网络“羊毛党”

IBM 雷佳音

目录

CONTENTS

01

对“羊毛党”的现状分析

02

数据分析软件介绍

03

模型部署平台介绍

04

模型算法介绍

05

建模实操练习

06

思维拓展

1

对网络“羊毛党”的现状分析

何为“羊毛党”？

何为“羊毛党”？

➤ “羊毛党”是关注与热衷于“薅羊毛”的群体，是指那些专门选择互联网公司的营销活动，以低成本甚至零成本换取高额奖励的人。



➤ 从最初的抢夺免费福利和优惠券，到近年来扎堆P2P网贷平台，再到聚集在电商平台上，“羊毛党”也逐渐从分散个体向组团集聚发展，形成了有组织、有规模、有分工的职业“羊毛党”，甚至发展出了完整、成熟的产业利益链。

➤ 尤其是随着电商之间的竞争越激烈，购物促销活动越来越多，“薅羊毛”的机会也会越来越多。

“羊毛党”有何特性？

➤ 智能化

购买大量的廉价智能手机用于注册账户，购买大量小额手机卡，使用短信集群收发的专业设备和自动化管理软件成功实现短信批量接收验证码的功能。

➤ 团伙化

“羊毛党”通过线上论坛，博客，社交群，或线下聚集等方式，形成群体。实现信息的交换及更高的利润收入。右图为“羊毛党”信息博客。



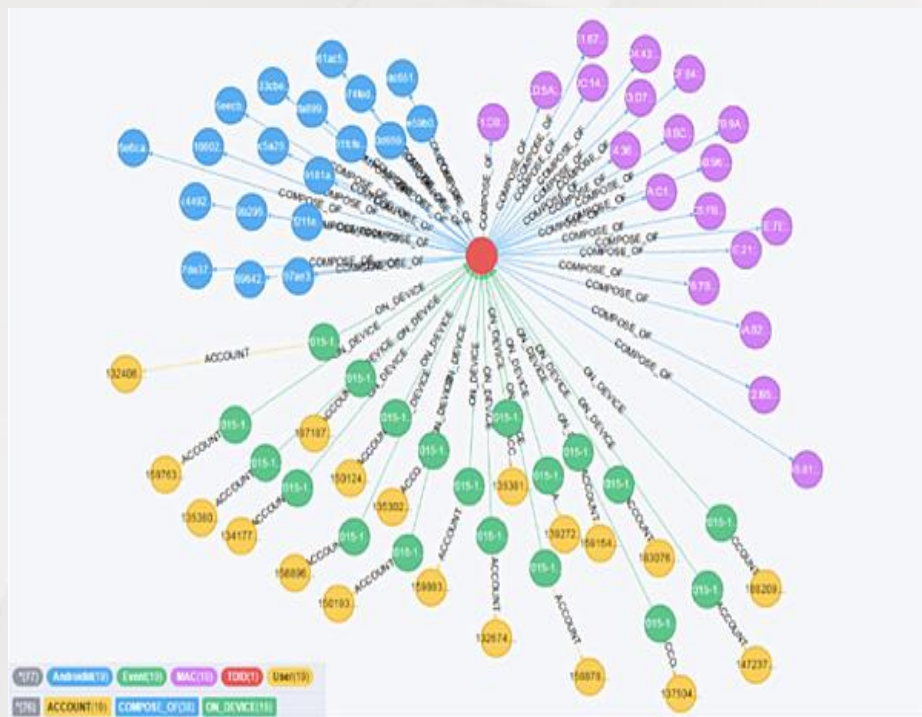
“羊毛党”分类

➤ 良性行为

个人自发行为，多为占便宜的心态。在“薅羊毛”的过程中，若发现这个平台可靠，进而会进行投资。这类“羊毛党”利于平台刷人气、完善现有规则等。

➤ 恶意敛财

个人通过不法渠道得到多张电话卡、身份证、银行卡等，在各个平台多次注册以多次享受同一优惠活动。挤占了正常用户参与活动的权益。右图描述同一用户用多个设备，多个账号，多次享受同一活动的情况。



甄别“羊毛党” 刻不容缓

- “羊毛党”已经渐渐从占便宜向非法盈利转移，从个人行为转为团伙集团活动。
- 恶意“羊毛团伙”对平台的活动运营成本损失很大，甚至导致平台直接被薅干，宣布提现困难，资金链断裂。
- 根据客户的特定消费记录，分析数据找到异常用户，推测作弊手段。在之后的促销活动中可以减少羊毛党带来的损失。
- 识别恶意“羊毛党”，对于商家正常盈利，维护市场平衡至关重要。

“羊毛党”行为痕迹

随着技术的发展，大量数据日志正在被大数据技术收集，处理，整合并应用，“羊毛党”的数据轨迹也逐渐完整与清晰。越来越多的商家意图通过保留的平台历史数据甄别恶意“羊毛党”。数据痕迹主要有以下几大类：

- 网络行为
- 交易行为
- 设备动态行为
- 手机的整体行为
- 平台行为

“羊毛党”行为痕迹



网络行为：

- 从抢夺免费福利和优惠券，到P2P网贷平台，再到聚集在电商平台上，“羊毛党”们以网络为主要阵地。
- 这些网络行为痕迹可以帮助我们甄别“羊毛党”，接入IP地址，Hostname，路由设备日志，运营商接入基站等，都可以留存大量的网络行为日志。
- 完整的网络日志可以形成一条“羊毛用户”网络路径，客观反映“羊毛党”的网络行为轨迹。

“羊毛党”行为痕迹



设备动态行为：

- 智能手机及其他智能手持设备内置的运动传感器收集设备的动态行为，如位置变化幅度，变化频次，变化规律等信息，从而通过数据计算判断设备的动态行为。
- 现常见商家及平台举营销活动，可将运动情况量化换取一定的报酬。例如：微信步数达到8000/每天，可以换取积分，礼品，现金等。但会有恶意“羊毛党”用户，利用工具，恶意刷运动记录等。
- 可通过设备的动态行为数据，甄别网络“羊毛党”。

“羊毛党”行为痕迹



平台行为：

- 被薅平台往往有很多的平台行为及过程，包括注册，绑卡，浏览，交易，提现。每个过程都会留下很多行为轨迹，而羊毛党特别是其中的机器“羊毛党”的行为轨迹更是有其特殊性。
- 这些特殊性可以帮助我们识别网络“羊毛党”，甚至对恶意网络“羊毛党”的意图进行分类。

“羊毛党”行为痕迹



交易行为：

- 商人无利不起早，“羊毛党”更是如此，“羊毛党”会对平台的产品做详细的对比分析，找出其中利益最大化的“薅羊毛”方案。
- 其交易的产品，交易金额和交易时间都是最佳化设计。比如，交易金额数非常小，但返还积分较高，礼品价值较高或返还现金较高。
- 通过“羊毛党”们往往总是追求利益最大化的薅羊毛方案这一特性，可以通过交易数据分析，甄别恶意“羊毛党”。

“羊毛党”行为痕迹



手机整体行为：

- 在移动互联网下，“羊毛党”的主要工具都是手机，每台手机上安装的互联网金融平台数，活跃时间，甚至于羊毛党对手机终端的偏好都可以留下一定的行为轨迹。
- 这些来自“羊毛党”的多方行为数据通过设备指纹，帐号ID，及更多的数据实体整合到一起形成一张“羊毛党”个人行为的数据图谱。
- 通过个人行为的数据图谱，可有效甄别出网络“羊毛党”。

2

数据分析软件介绍

IBM SPSS Modeler

IBM SPSS Modeler

SPSS Modeler 是领先的**可视化数据科学**和**机器学习**解决方案。

<https://www.ibm.com/cn-zh/products/spss-modeler>

搜索

SPSS Modeler 详细信息 定价 支持 资源 

免费产品展示

IBM SPSS Modeler

借助拖放式数据科学工具，推动提升 ROI。该产品可作为 IBM Watson Studio 的一部分提供，也可作为独立订购或永久产品提供。

SPSS 企业版免费试用

SPSS 学生版免费试用

查看定价并购买



Watson Studio 即将推出桌面版 → 体验测试版

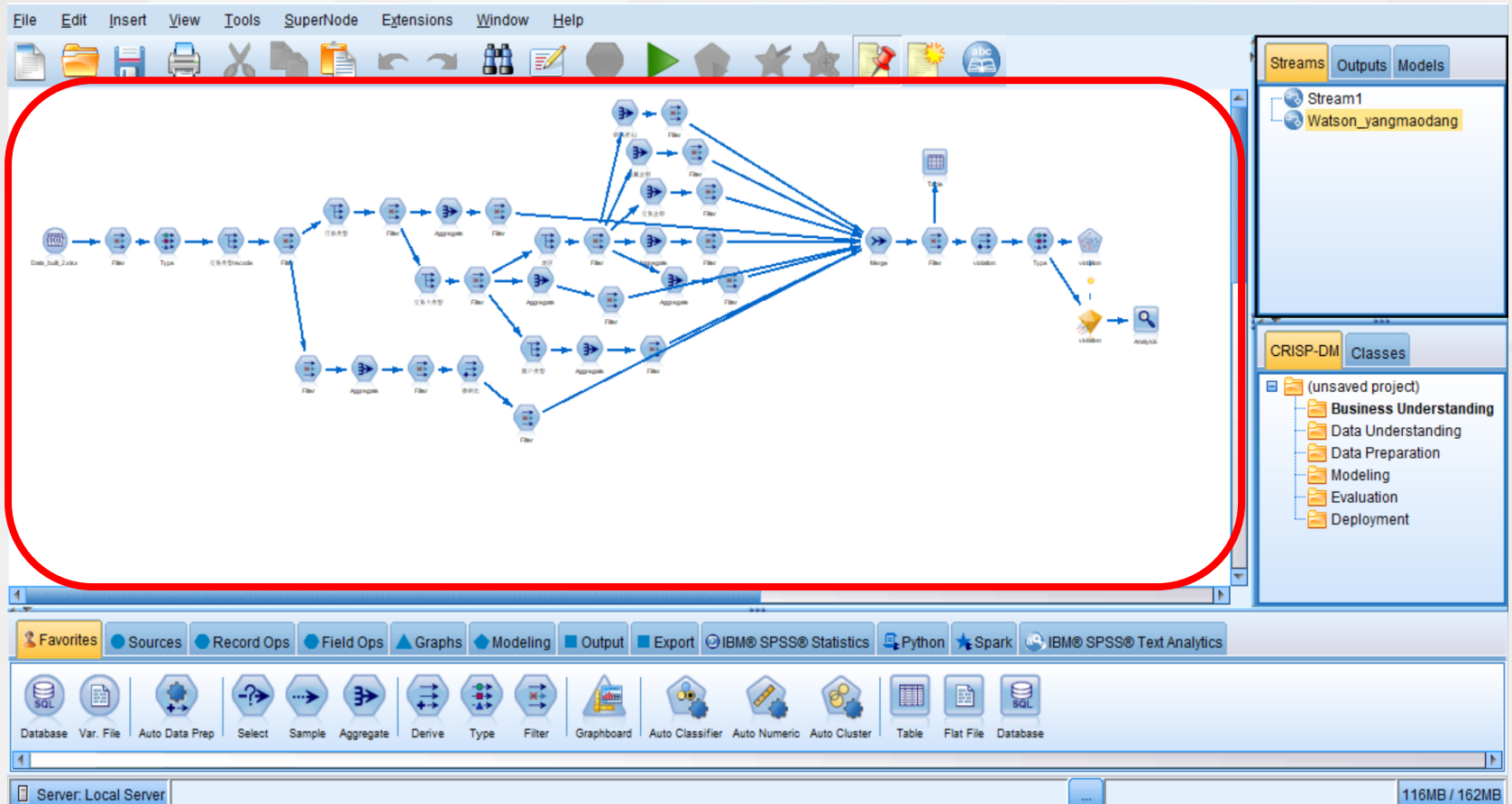
IBM SPSS Modeler 新增功能

联系 IBM

IBM SPSS Modeler

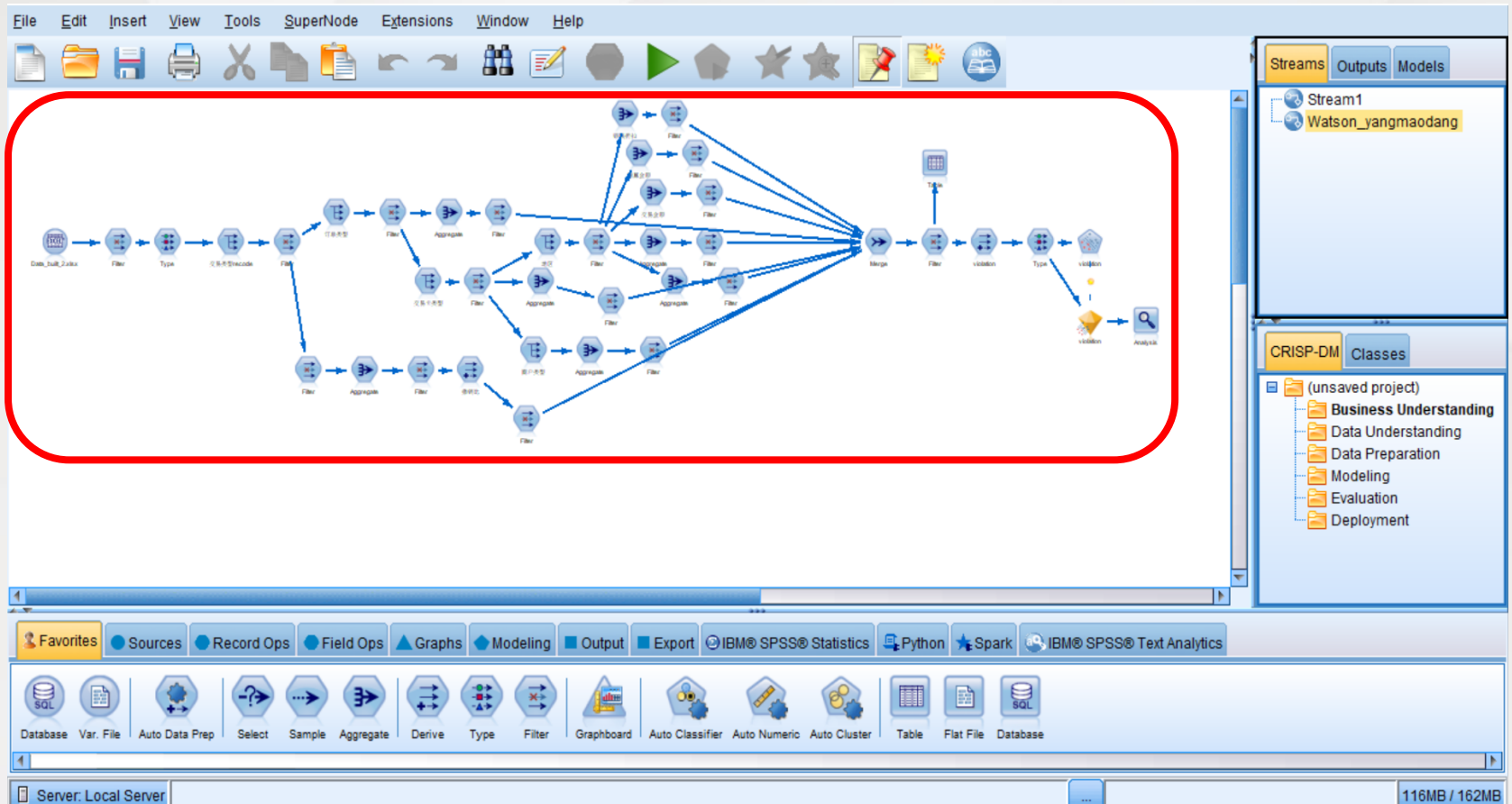
- 加快数据研究员执行操作任务的速度，从而帮助企业加速实现价值并获得预期的成果。全球领先的组织依靠 IBM 的产品和服务进行数据准备和发现、预测分析、模型管理和部署以及机器学习，通过数据资产获得实际收益
- SPSS Modeler 支持企业利用数据资产和现代应用程序，以及准备就绪且立即可用的完整算法和模型。它适用于混合环境，可以满足监管和安全需求 — 通过 IBM Watson® Studio 提供。SPSS Modeler 可帮助您：
 - 充分利用基于开源的创新，包括 R 或 Python。
 - 支持拥有不同技能（编程和可视化）的数据研究员。
 - 开发混合方法 — 在本地、公共云或私有云中。
 - 从小规模开始，扩展至企业级监管方法。

IBM SPSS Modeler



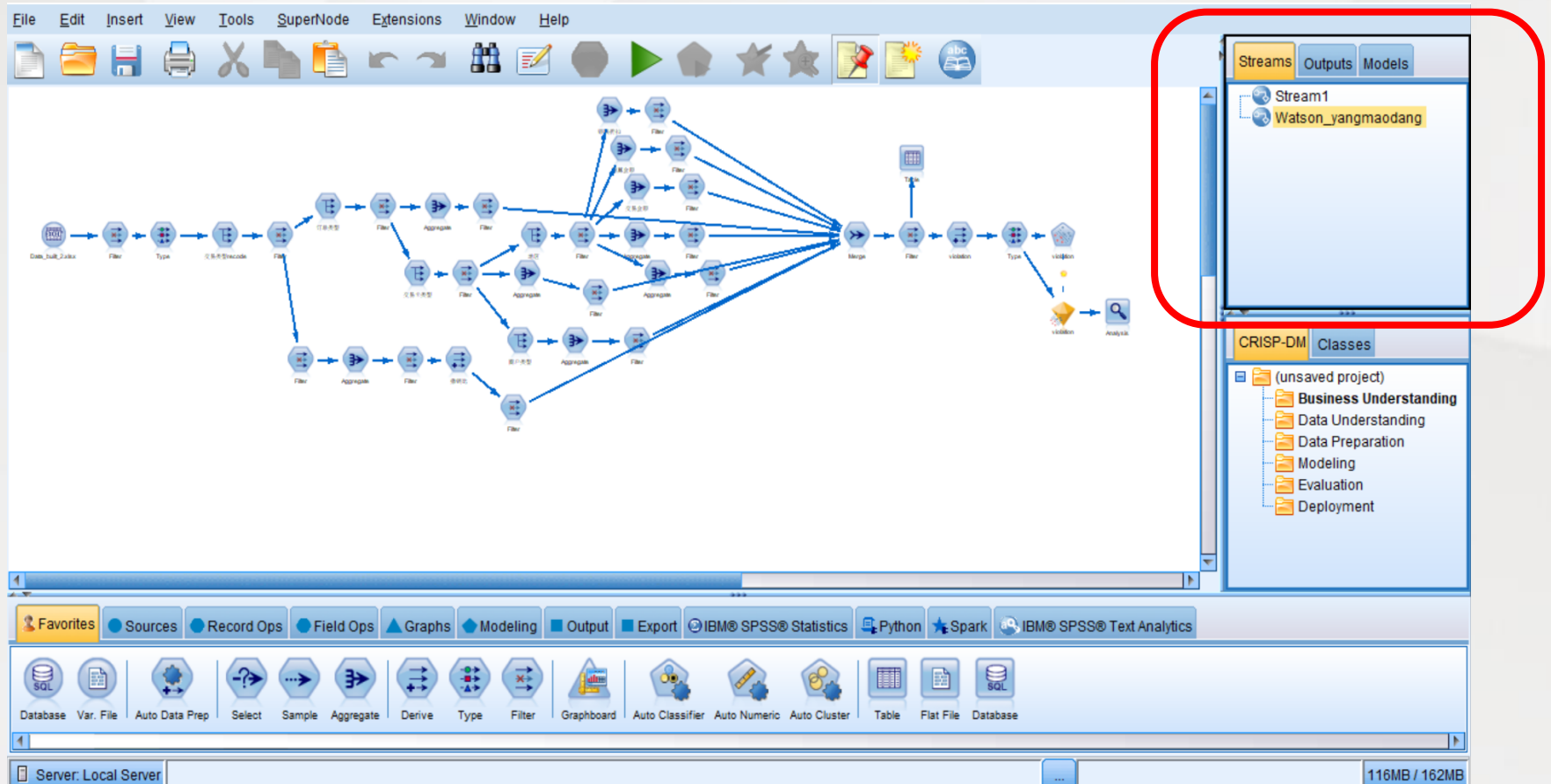
流工作区是 IBM SPSS Modeler 窗口的最大区域，也是您构建和操纵数据流的位置。

IBM SPSS Modeler



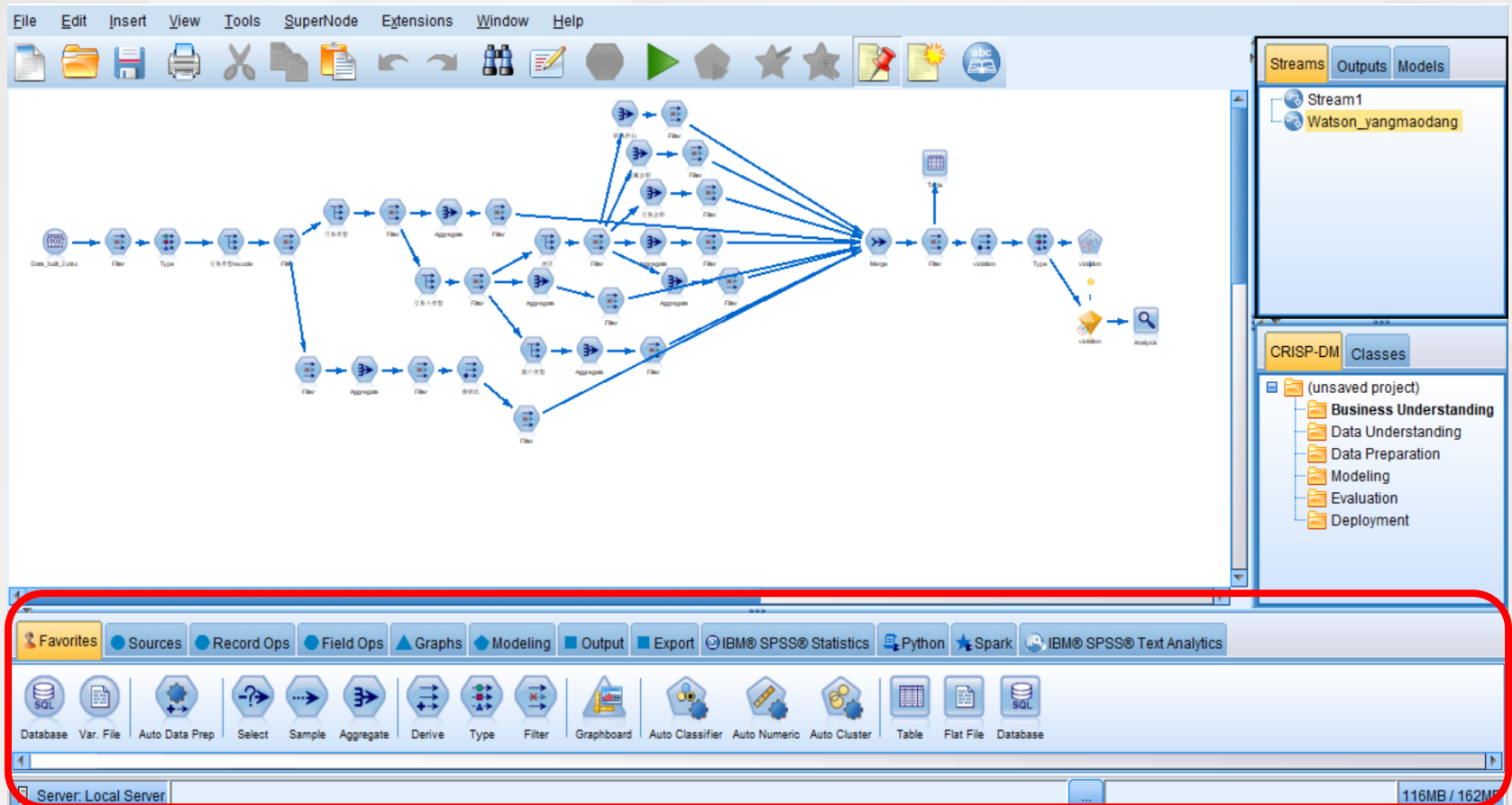
流是在界面的主工作区中通过绘制与业务相关的数据操作图来创建的。每个操作都用一个图标或节点表示，这些节点通过流链接在一起，流表示数据在各个操作之间的流动。

IBM SPSS Modeler



在 IBM SPSS Modeler 中，可以在同一流工作区或通过打开新的流工作区来一次处理多个流。会话期间，流存储在 IBM SPSS Modeler 窗口右上角的“流”管理器中。

IBM SPSS Modeler



数据和建模工具可从节点选用板获取。双击图标将节点添加到画布中，连接各个图标以创建一个表示数据流动的流。

IBM SPSS Modeler 介绍

- **记录选项节点：**对数据记录执行操作，例如选择、合并和附加。
- **字段选项节点：**对数据字段执行操作，例如过滤、派生新字段和确定给定字段的测量级别。
- **图形节点：**以图形方式显示建模前后的数据。包括散点图、直方图、网络节点和评估图表。
- **建模节点：**使用 SPSS Modeler 中提供的建模算法，例如神经网络、决策树、集群算法和数据序列等。
- **数据建模节点：**使用 Microsoft SQL Server、IBM DB2 以及 Oracle 和 Netezza 数据库中提供的建模算法。
- **输出节点：**为数据、图表和模型结果生成可以在 SPSS Modeler 中查看的各种输出。
- **导出节点：**生成可以在外部应用程序中查看的各种输出。

3

模型部署平台介绍

IBM Watson Studio

IBM Watson Studio

Watson Studio是一个为多个用户提供分析工具，共同进行数据分析，解决业务问题的集成分析平台。

<https://www.ibm.com/cloud/watson-studio>

IBM | Cloud | Why IBM | Cloud Paks | Products | Solutions | Garage | Pricing | More ▾

Search [Cloud sign-up/log-in](#)

Watson Studio | Details | Pricing | Solutions | AutoAI | Resources ▾ [Sign into IBM Watson Studio Cloud](#)

IBM Watson Studio

Simplify and scale data science to predict and optimize your business outcomes

[Start on cloud for free](#) [Try on desktop for free](#)

→ [Explore Watson Studio Local product tour](#)

IBM Watson® Studio Desktop is available at a special [View pricing and buy](#)

[Let's talk](#)

IBM Watson Studio

Watson Studio支持端对端的AI 人工智能工作流程。

Watson Studio

Supporting the end-to-end AI workflow

Connect &
Access Data

Connect and discover content from multiple data sources in the cloud or on premises. Bring **structured** and **unstructured** data to one toolkit.

Search and Find
Relevant Data

Find data (structured, unstructured) and AI assets (e.g., ML/DL models, notebooks, Watson Data Kits) in the **Knowledge Catalog** with intelligent search and giving the right access to the right users.

Prepare Data
for Analysis

Clean and prepare your data with **Data Refinery**, a tool to create data preparation pipelines visually. Use popular open source libraries to prepare unstructured data.

Build and Train
ML/DL Models

Democratize the creation of ML and DL models. Design your AI models **programmatically** or **visually** with the most popular **open source** and IBM ML/DL frameworks or leverage transfer learning on **pre-trained** models using **Watson tools** to adapt to your business domain. Train at scale on **GPUs** and **distributed** compute

Deploy Models

Deploy your models easily anywhere and have them **scale automatically** for online, batch or streaming use cases

Monitor, Analyze
and Manage

Monitor the performance of the models in production and trigger automatic retraining and redeployment of models. Build **Enterprise Trust** with Bias Detection, Mitigation Model **Robustness** and Testing Service Model **Security**.

IBM Watson Studio

Watson Studio可以让具有不同技能的团队进行合作。

Watson Studio

Built for AI teams – enabling team productivity and collaboration



Tanya
Domain Expert

Her Job:

To transfer knowledge to Watson for a successful user experience.

What she does:

- Range of domain knowledge and uses that to teach Watson and develop a custom models
- As Tanya gains more experience she optimizes her knowledge to teach Watson to design better end-user experiences.

Sometimes known as:

Subject matter expert, content strategist.



Mike
Data Scientist

His Job:

Transform data into knowledge for solving business problems.

What he does:

- Runs experiments to build custom models that solve business problems.
- Use techniques such as Machine Learning or Deep Learning and works with Tanya to validate success of trained models.

Sometimes known as:

ML/DL engineer, Modeler, Data Miner



Ed
Data Engineer

His Job:

Architects how data is organized and ensures operability

What he does:

- Builds data infrastructure and ETL pipelines. Works with Spark, Hadoop, and HDFS.
- Works with data scientist to transform research models into production quality systems.

Sometimes known as:

Data infrastructure engineer



Deb
The Developer

Her Job:

Builds AI application that meet the requirements of the business.

What she does:

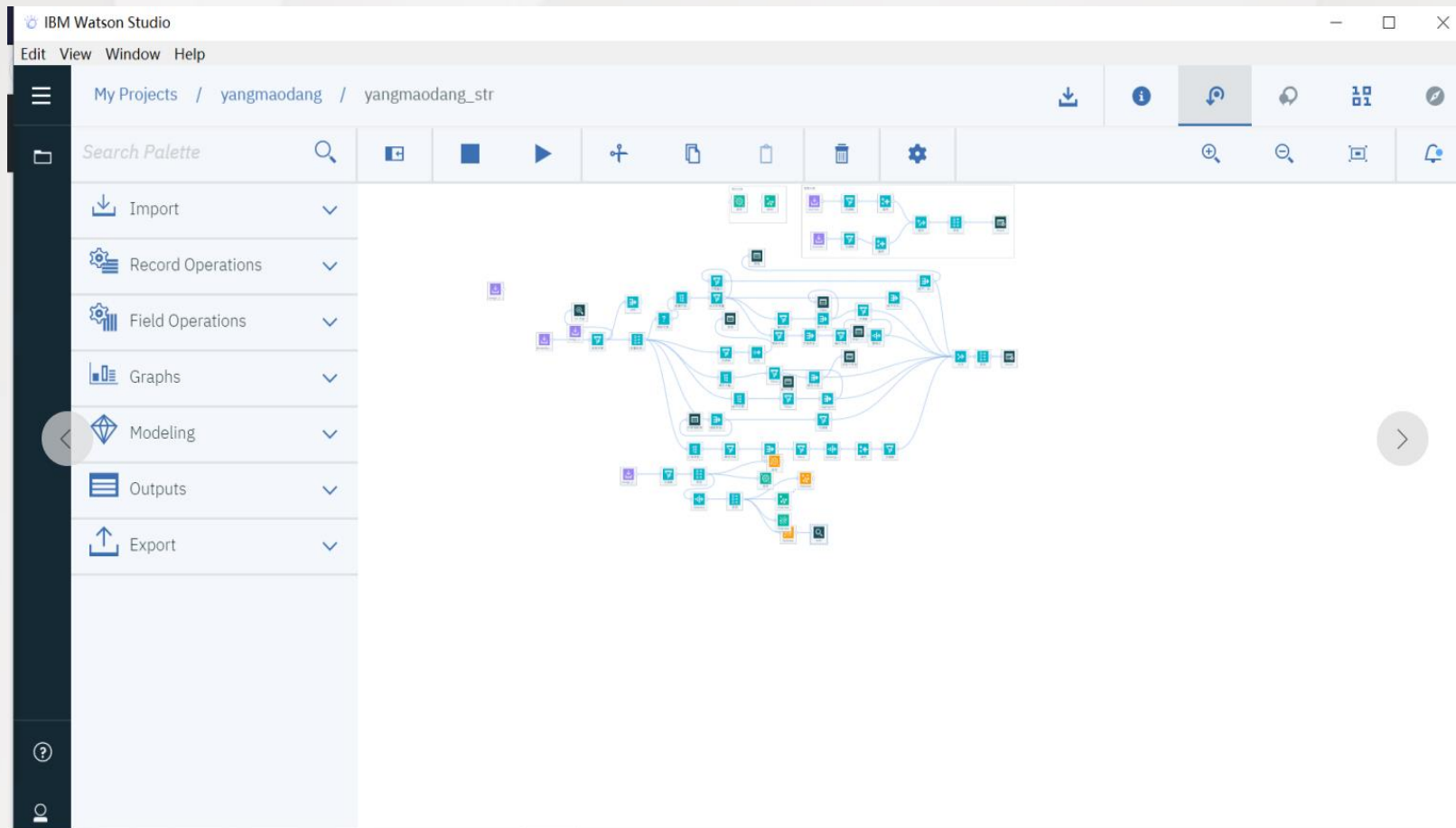
- Starts PoCs which includes gathering content, dialog building and model training
- Focus is on app building for the team or company to use. Will handle ML Ops as needed

Sometimes known as:

Front-end, back-end, full stack, mobile or low-code developer

IBM Watson Studio

IBM Watson Studio 中的SPSS Modeler 界面



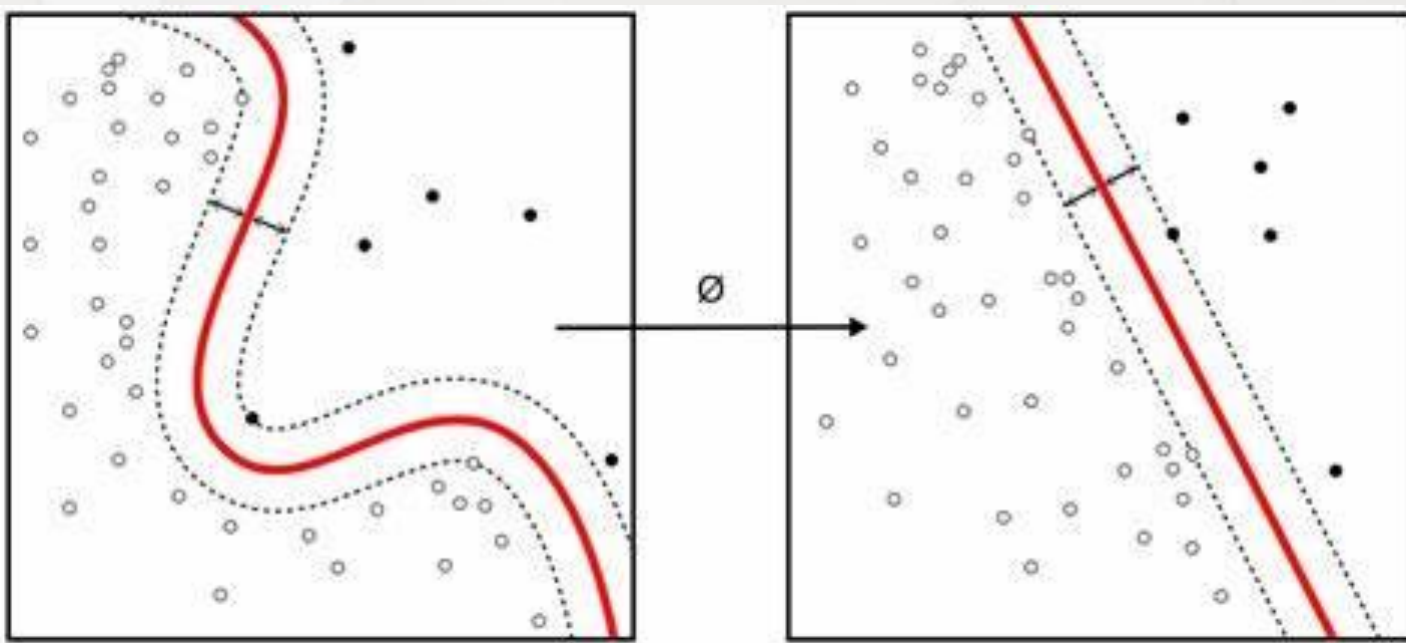
4

模型算法介绍

支持向量机 SVM

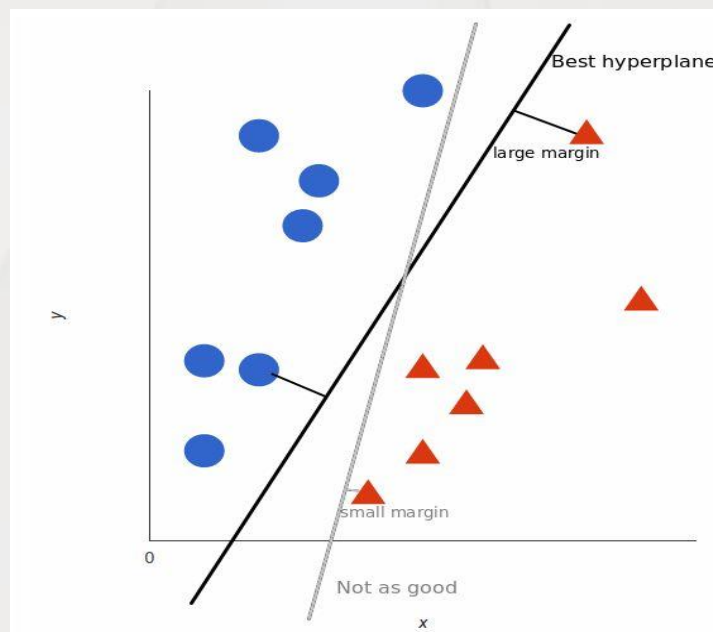
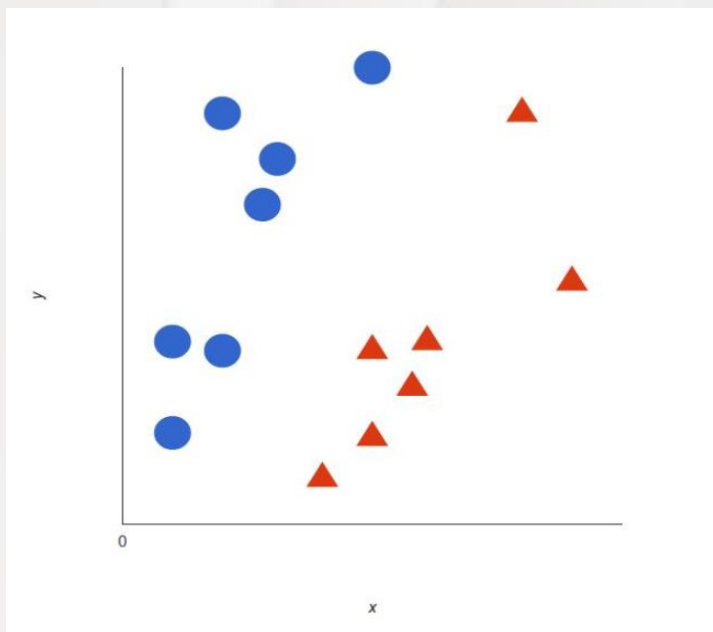
支持向量机 SVM

- 支持向量机 (Support Vector Machine, SVM) 是一类按监督学习方式对数据进行二元分类的广义线性分类器, 其它的决策边界是对样本求解最大边距超平面。
- SVM可以通过核方法进行非线性分类。



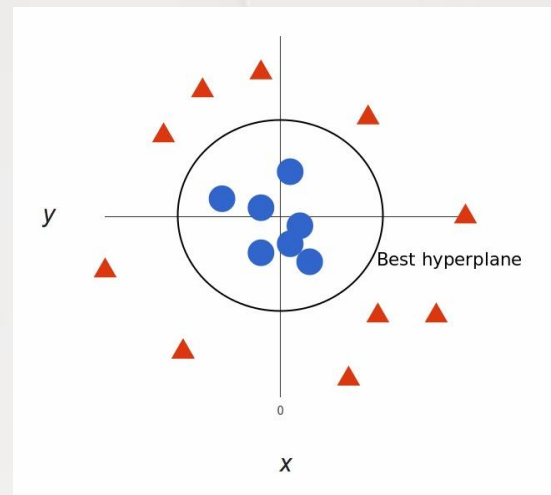
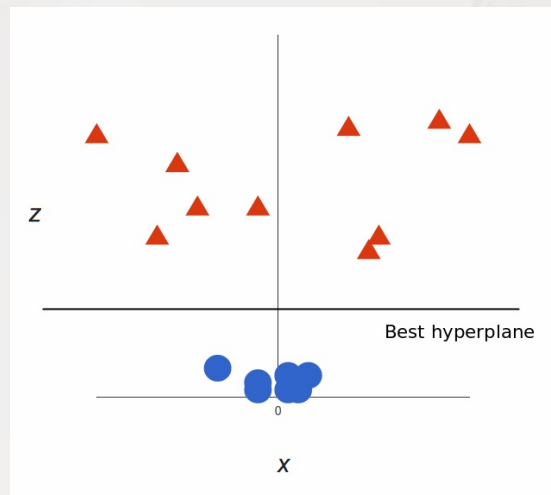
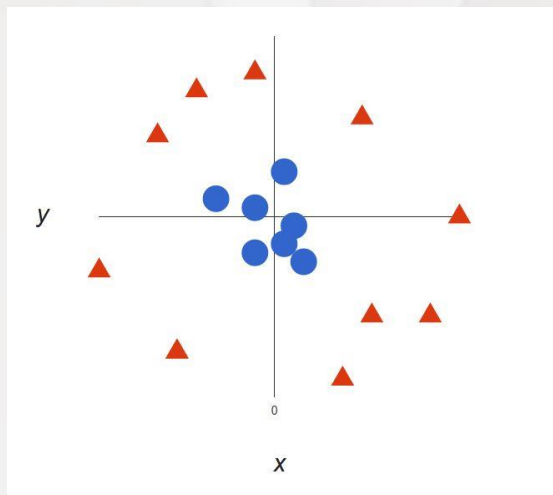
支持向量机 SVM

让我们想象两个类别：红色和蓝色，我们的数据有两个特征 x , y 。我们想要一个分类器，给定一对坐标 (x, y) ，输出仅限于红色和蓝色。我们将已标记的训练数据列在下图中：支持向量机会接收这些数据点，并输出一个超平面将这两类分隔开。最好的超平面（在本例中是一条线）对每个类别最近的元素距离最远。



支持向量机SVM

更多的时候，你无法找出一个线性决策边界（一条直线分开两个类别）。这时我们引入第三个维度。加入维度 z ，并以直观的方式出现： $z = x^2 + y^2$ 。于是有了了一个三维空间，我们可以在三维空间找到一个线性决策边界。现在我们处于三维空间，超平面是 z 某个刻度上（比如 $z=1$ ）一个平行于 x 轴的平面。它在二维上的投影是一个圆。



5

建模实操练习

利用SPSS Modeler 建模

“羊毛党” 分析实例

数据清洗

对数据进行重新审查和校验。删除重复信息、纠正存在的错误，并提供数据一致性。

数据探索

运用统计方法对数据进行探索，发现数据内部规律。根据具体情况，对数据进行转换。

数据转换

为了达到模型的数据模型的要求，需要对数据转换，包括归一化，标准化等。

构建模型

尝试多种模型，通过后续的模型评估，对模型进行优化，调整，以寻求最有解。

模型评估

对模型的精度，准确度，效率和通用性进行评估。模型是否回答了最初的业务问题。

6

更多思路

根据实际情况具体分析

甄别“羊毛党”更多思路

我们还可以从其他方面的数据痕迹进行分析：

- 网络行为
- 交易行为
- 设备动态行为
- 手机的整体行为
- 平台行为

对于其他数据，我们需要具体情况具体分析，找到最合适的解决方案。

Q&A

THANK