**数据挖掘系统**

**可行性研究报告**

1. **引言**
   1. **编写目的：**

研究此次软件工程项目是否具有研发的必要，是否可以实现，是否具有市场前景，便于项目投资者进行评估，是项目投资的主要依据。

项目进行工程设计、设备订货、施工准备等基本建设前期工作的依据。是编制设计文件、进行建设准备工作的主要依据。

为项目拟采用的新技术提供辩证理论依据。

可行性研究的目的是对问题进行研究，以最小的代价在最短的时间内确定问题是否可解。经过对此项目进行详细调查研究，初拟系统的可行性报告，对软件开发中将要面临的问题及其解决方案进行初步设计及合理安排，明确开发风险及其带来的经济效益。

本报告经审核后，交由项目经理审查。

* 1. **背景：**

开发软件名称：数据挖掘系统

项目任务提出者:：范渐臻

项目开发者：范渐臻及其团队（陈靖舸 苏锐 张亮亮）

用户：有需求的客户

（1）数据挖掘的技术背景

数据挖掘是八十年代投资人工智能研究项目失败后人工智能转入实际应用时提出的。它是一个新兴的面向商业应用的人工智能研究。选择数据挖掘这一术语表明了与统计、精算、长期从事预言模型的经济学家之间没有技术的重叠。数据挖掘技术包括三个主要部分——算法和技术、数据、建模能力。

（2）数据挖掘的商业背景

数据挖掘首先是需要商业环境中收集了大量数据并要求挖掘的知识是有价值的。

对商业而言有价值主要表现在三个方面——降低开销、提高收入、增加股票价格。

在商业运营中数据挖掘主要用作以下四种工具

1)数据挖掘作为研究工具Research

2)数据挖掘提高过程控制ProcessImprovement

3)数据挖掘作为市场营销工具Marketing

4)数据挖掘作为客户关系管理CRM工具(CustomerRelationshipManagement)

（3）相关技术术语

1）机器学习

机器学习是计算机科学和人工智能发展的产物

机器学习分为两种学习方式、自组织学习（如神经网络)和从例子中归纳出规则（如决策树）。

2）统计

统计本包括预言算法、回归、抽样、基于经验的设计等，现在也开始支 持数据挖掘。

3）决策支持系统

4）数据仓库

5）OLAP：联机分析处理、DataMart：数据集市、多维数据库等

将数据仓库、OLAP、数据挖掘等技术融合在一起即构成企业决策分析环境。

* 1. **定义：**

轻量级的数据挖掘系统：

对数据（sql或是csv文件）进行分类，聚类，关联，推荐等操作，再提供可视化的数据显示界面，提供用户数据爬取的接口。初始版本为本地版本，后期可能会移植网站服务。

* 1. **参考资料：**

张海藩，吕云翔.实用软件工程[M].北京:人民邮电出版社，2015

**2 可行性研究的前提**

**2.1 要求：**

保证小规模的数据得到挖掘和训练，以方便初学者学习和小规模数据量的分析工作，便于日常使用。

**2.2 目标**

现在非常火热的自媒体或是视频制作者对简单易操作的数据分析和可视化工具的需求也十分迫切，便于论证他们的视频内容（数据来源于自己收集，而非官方公布的结果）。或者是炒股的股民将交易的记录进行分析。

设计一个主要针对sql、csv数据文件进行挖掘分析系统，提供用户可视化的操作界面，只需要导入小规模数据文件（不超过万条），再选择相应的分析类型（如聚类，分类，关联等）和对应的参数后，就可以得到分析模型和可视化的分析结果。

分析的模型会存储在本地，待新的数据来临时，只需要调用存储的模型就可以判断新数据的类型。新数据的导入可以选择手动输入，或是导入新的数据文件，但是要注意格式一致（系统会有提示）。

我们也会给用户提供数据爬取的接口，爬取后的数据会存入本地，以供用户使用。

**2.3 条件、假定和限制**

新数据的导入可以选择手动输入，或是导入新的数据文件，但是要注意格式一致（系统会有提示）。

**3 对现有系统的分析**

**3.1 处理流程和数据流程**

****

**3.2 工作负荷**

**3.3 费用开支**

由于本项目短期并不

**3.4 人员**

范渐臻 安徽大学本科在读 软件项目主程序员

苏锐 安徽大学本科在读 软件文档撰写者

张亮亮 安徽大学本科在读 软件项目副程序员，测试人员

陈靖舸 安徽大学本科在读 算法和数学研究员

**3.5 设备**

介于本次项目是本地的软件开发，只需要在个人电脑上安装开发环境即可。初步项目没有涉及到网络服务，也没有设置服务器的必要。

**3.6 局限性**

1）由于如果处理大型数据流，会导致计算机算力不足

2）主要是提供本地的数据分析工作，还不能投入市场使用

**4 所建议的系统**

**4.1 对所建议系统的说明**

提供用户一个轻量级数据挖掘系统，对数据（sql或是csv文件）进行分类，聚类，关联，推荐等操作，再提供可视化的数据显示界面，提供用户数据爬取的接口。初始版本为本地版本，后期可能会移植网站服务。

总体而言，系统的初步分为四个模块：

①数据分析模块：旨在用已有的数据训练处模型；

②模型存储模块：将训练好的模型进行存储，以便新数据来临时的归类；

③数据爬取模块：为用户提供在网上爬取数据的模块，可以作为训练的原型。

④可视化操作界面模块：为用户提供简单方面的操作体验，只需要导入文件和输入参数就可进行操作。

**4.2 处理流程和数据流程**

（1）用户根据可视化的操作界面，只需要导入小规模数据文件（不超过万条）

（2）选择相应的分析类型（如聚类，分类，关联等）和对应的参数后，就可以得到分析模型和可视化的分析结果。

（3）分析的模型会存储在本地，待新的数据来临时，只需要调用存储的模型就可以判断新数据的类型。

新数据的导入可以选择手动输入，或是导入新的数据文件，但是要注意格式一致（系统会有提示）。

我们也会给用户提供数据爬取的接口，爬取后的数据会存入本地，以供用户使用。

**4.3 改进之处**

未来升级网络服务时，会租赁云服务器、移植网站服务。

**4.4 影响**

**4.4.1 对设备的影响**

软件系统完全兼容原有设备，不需要更换或改造设备。

**4.4.2 对软件的影响**

该浏览器兼容现存应用软件和相关支持软件，无需对这些软件进行修改和补充。

**4.4.3 对用户单位机构的影响**

简单易用，不需要用户单位机构设置专业人员进行管理和维护

**4.4.4 对系统运行过程的影响**

对系统运行过程的影响如下。

①用户操作规程与原系统基本一致。

②运行中心与用户通过该软件实现联系。

③用户登录及浏览数据存人服务器端的相应数据库

**4.4.5 对开发的影响**

对开发的影响如下。

①需要雇佣一些开发人员进行产品开发。

②需要租借开发人员办公场所。

③需要一定数量的计算机进行开发。

④需要建立数据库以提供技术支持。

**4.4.6 对地点和设施的影响**

无需额外的使用场所，不需改造现有设施。

**4.4.7 对经费开支的影响**

该浏览器开发难度较大，开支项主要有开发人员的设备、使用的计算机的购买或租赁费用、系统及数据库建设和维护费用。。

**4.5 局限性**

由于刚刚涉足数据挖掘以及软件工程开发，支持的扩展性功能可能比较少，不容易实现丰富的扩展性功能。随着使用人数的增加，其支持的功能会逐渐增多，该问题会逐渐改善**。**

**4.6 技术条件方面的可行性**

我组在之前的课程中接触过机器学习的相关概念，对相关的知识信息有着继续学习的浓厚兴趣，相关的数学基础我组成员也都具备，一般的经典算法我组成员也都掌握。

在开发平台方面，我组成员有过开发python项目的经验，对相关的数据库模块，数据可视化模块以及用户UI模块都有较为熟悉的开发经验。

因此，即使本次项目的开发工作十分复杂，我组成员也都做好了准备。

**5社会因素方面的可行性**

**5.1 法律方面的可行性**

（1）使用的软件开发工具都是合法，正确的使用。

（2）得到的数据仅用于学习，不用于商业 **。**

**5.2 使用方面的可行性**

我们会尽可能的简化操作，实现用户界面的可视化易用性。

就硬件条件而言，各种外围设备、计算机设备的性能能够满足系统的开发，并充分发挥其效应。

对软件条件来说，技术人员精通开发所需的各种软件。

**6结论**

通过对此软件系统进行的各方面的可行性分析，可以得出以下结论。

(1)对于数据挖掘以及数据分析对于生活的帮助是很大的，用户量也会比较大，用户角色也比较丰富。

(2)数据分析师正是社会亟需的人才，我们的产品能够很好地解决这方面的需求。

(3)在产品投入使用后，使用合理可以帮助我们更好做一些未来的规划

综上所述，全功能的通用型网络浏览器是一款能够 提高用户浏览体验的软件系统，可以立即

进行此软件系统的开发。。