<https://github.com/nd009/capstone/tree/master/Rossmann_Store_Sales>

https://www.kaggle.com/c/rossmann-store-sales/data kaggle连接

在此毕业设计中，您将运用您在此纳米学位计划中学到的机器学习的算法和方法选择一个你感兴趣的问题来解决。首先你需要定义这个你想要解决的问题，以及它的衡量标准。其次，你需要通过数据可视化和数据挖掘分析这个问题，对适于解决这个问题的算法和特征有更好的了解。接着，你需要实现你的算法和测试标准。你需要记录下前处理、改善、后处理的整个过程。然后，你需要收集这些模型的表现结果，把重要的部分可视化，来验证和证明这些结果。最后，你要在你的结果基础上得出结论，说明你的实现是否真的解决了这个问题

希望通过设计这样一个项目，让你撰写一篇完善的、端到端的，针对你感兴趣问题的解决方案的报告。无论是研发新技术、从之前的技术中作改进，学会完善的记录你的流程，是一项对于结果验证，重现，不可或缺的技能

通过完成这个项目，你将学到：

如何研究和探索一个你感兴趣的真实世界的问题。

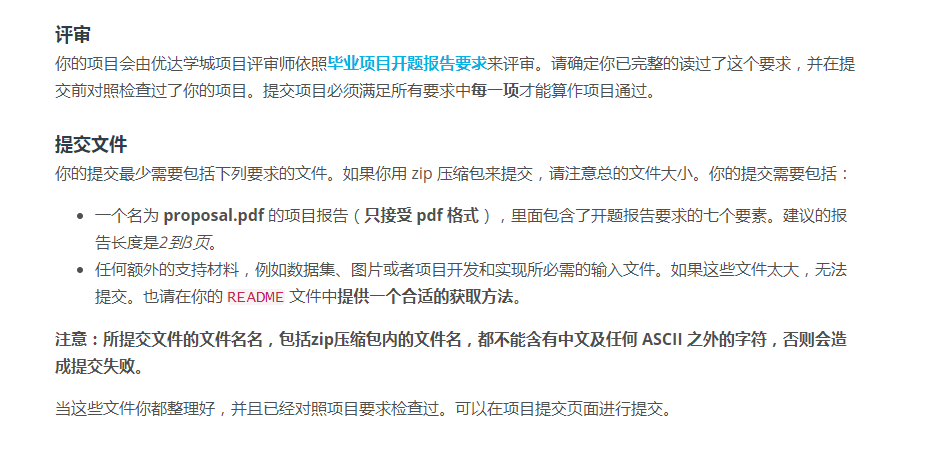
如何准确的应用一个特定的机器学习算法和技术。

如何对数据进行分析和可视化，以及如何验证结果。

如何把你的工作记录下来并形成报告。

无论你最终选择了哪个领域的问题，你都需要用5个主要步骤来完成这个项目，并把步骤记录下来。从问题的定义，到做出一个解决方案。每个步骤在这整个过程中都扮演了非常重要的角色。在你完成项目的过程中，记得要为你的项目报告打草稿，这也是你项目提交的最重要的方面。

如果没有按照这个要求写，你的报告可能不会得到充分的审阅。我们特地为你准备了一个开题报告模版，帮助你了<https://github.com/nd009/capstone/blob/master/capstone_proposal_template.md>





标准

符合要求

项目背景

学生简要的描述了毕业项目的背景（历史）信息。明确的陈述了为什么这个问题应该并且能够得到解决。引用到的学术研究都应该注明出处。建议学生谈一下选择这个领域中该特定问题的原因（可选）。

问题描述

学生清晰的描述了需要解决的问题。问题被明确的定义出来并且至少有一个可能的解决办法。此外，还要求这个问题可以被量化，被衡量以及可重现的。

数据或输入

对项目中用到的数据集或输入有详尽的描述。包括如何获取这些数据集或者输入，他们有何特点。 还应当说明数据集或输入在项目中如何使用，用法对解决这个特定问题是否得当。

解决方法描述

学生清晰的描述了解决问题的方法。这个解决方法能够配合数据集或者输入解决这个问题。 此外，还要求这个解决方法可以被量化，被衡量以及可重现的。

评估标准

学生提出了至少一个用于量化基准模型和解决方案的的评估标准。这个评估标准对于问题本身、数据集以及解决方案来说都是合适的。

基准模型

学生提供了相关问题领域解决方案的基准模型。理想情况下，学生的基准模型包括已有方法的描述，该领域内已知的信息，使得学生的解决方案可以和该基准模型作客观的对比。该基准模型定义清晰且可衡量。

项目设计

学生总结了一个针对问题解决方案的实施理论流程。探讨了计划采取的策略，对数据需要进行哪些分析，考虑哪些算法。这些流程和探讨符合该问题的特点。我们鼓励把数据简单可视化，加入一些对解释有帮助的伪代码及图表。

报告呈现

整个开题报告结构清晰，易读易懂。每一个部分清楚简洁。几乎没有错别字或语法错误。所有资源及引用都有注明出处。