文档编号：2018031101

版本号：V 1.0

{客户贷款预测}

项目说明书

{XX金融服务有限公司}

2018年3月

目录

**1.概述 …………………………………………………1**

1.1编写目的 ……………………………………………………2

1.2项目背景 ……………………………………………………5

1.3名词定义 ……………………………………………………6

1.4参考资料 ……………………………………………………10

**2.数据库数据说明**

2.1数据库基本说明 ……………………………………………………15

2.2 数据逻辑结构 ……………………………………………………17

**3.业务数据分析**

3.1 衍生变量 ……………………………………………………18

3.2 变量关联性分析 ……………………………………………………23

**4.模型概述**

4.1模型说明 ……………………………………………………28

4.2模型要求 ……………………………………………………35

1.概述

**1.1编写目的**

项目说明文档包含三个部分，1.项目概述 2.数据说明 3.模型说明。此文档供技术人员参考使用，请仔细阅读。第一部分说明了业务需求，第二部分展示了我司现有的相关数据，第三部分对模型的构建做出了明确指导。

**1.2 项目背景**

随着我司近期金融板块的不断发展，贷款产品也需求也日益增加。为了扩大生产利益，同时更好的服务客户，现需开发贷款需求预测模型， 以便配合产品部门进行有目标有规划的产品推荐。

**1.3 名词定义**

注：只对易混淆名词进行解释，一般名词不做特殊解释

active\_date：激活日期，非用户注册日期，是用户注册之后，首次消费的日期

limit : 初始额度，即用户注册之日，首次使用测评获得的额度

discount：优惠金额，即用户在各节日获取的免减卷或赠金活动的折合计算

param：页面参数，某个产品下的用户点击行为记录

2.数据库数据说明

2.1 数据库基本说明

我司使用mysql数据库，相关数据使用csv导出，请自行导入本地数据库。

相关数据表格总计5张，总计数据量约150w条，占用内存大小约1G。

本地电脑需求：1.内存4g以上 2.本地硬盘容量不少于30G

2.2 数据逻辑结构

数据表格及名词解释如下：



3.业务数据分析

3.1 衍生字段

给予t\_order t\_click t\_loan三表做衍生变量，

衍生变量字段名称 和 计算公式 见附件《features\_list》

3.2 变量关联性分析

根据附件生成衍生变量后，请完成这一部分

4.模型概述

4.1 模型说明

此模型本质为回归模型，意在以消费者前一段时间的行为和消费者自身特征来预测其下一个时间段的贷款额，此模型可以从两个方向着手。

1、 直接建立回归模型，通过已有字段和衍生字段，预测消费者贷款。

2、首先判断消费者是否会贷款，建立分类模型，之后对会贷款的消费建立回归模型，预测其具体的贷款额。

4.2 模型要求

请使用可解释性的算法来建模，建模结果必须可解释，使用字段必须可描述其重要性。