项目阶段一

1. 项目简介：
2. 、项目背景：

手机app“懂球帝”是近几年上市的一款足球新闻app，上市以来，使用人数逐年增加，受到广泛球迷的好评和热爱。如今“懂球帝”已然成为中国最受欢迎、使用量最大的足球app之一，其新闻热度、评论区可以一定程度上可以代表中国球迷意见。而且懂球帝评论管理十分严格，鲜有没有意义的评论或者谩骂，因此有很大的挖掘意义。

巴萨、皇马是国际足坛上两大天王山之战级别的豪门，双方在国际上的拥垒数以亿计，在中国也有很大的市场，国家德比全球同时在线观看人数达到12亿之多可见其影响力。在习总书记的支持下，我国足球事业蒸蒸日上，那么也不可忽略这两大豪门对中国的影响。虽然每年都有官方数据统计，但可信程度受到质疑，本组作为巴萨球迷，想对两队支持者人数比例一探究竟。

1. 、项目方案：

通过爬虫从“懂球帝”官网www.dongqiudi.com开始爬出近期新闻，格式www.dongqiudi.com/article/\*，从而得出新闻中所有评论用户（据使用懂球帝2年多经验可知，用户命名时多用主队名、主队教练名、主队球星名、主队球场名等命名）。继而得出活跃的“懂球帝”们所支持的主队，从中统计出巴萨、皇马的支持者（之所以挑选评论区的用户，是因为活跃参与评论的多半是真球迷）。得出报表，分析出结论。存在的不足之处主要在于爬取到的数据以用户名为关键词，并不能涵盖全部支持者，但所取数据很有代表性，可以推得整体比例。

1. 、项目使用技术：

WebMagic