资本市场金融科技创新试点（北京）项目公示表

填报时间： 2021 年 7 月 21 日

一、项目概览	1.1 项目编号	BJ-SD-202116
	1.2 项目名称	基于联邦学习技术的强监管营销模型的探索
	1.3 项目类型	业务辅助类
	1.4 项目简介	机器学习在建信基金各方面业务中均发挥了重要作用，创造了可观的业务价值。建信基金与某大型国有商业银行联合开展“基于联邦学习技术的强监管营销模型的探索”项目，在对用户数据安全日益严格的监管要求下，探索机器学习中的“联邦学习” 前沿技术在精准营销业务中的应用。 该技术通过加密机制在联邦系统内进行参数交换的方式建立一个共有预测模型，从而在各数据参与方不泄露客户隐私和底层数据的前提下，利用更丰富的数据提升预测模型的精准度。通过联邦学习技术得到的模型对客户进行精准营销，可以提高营销效率，降低营销成本。在 2020 年进行的短信营销活动中，根据联邦学习模型筛选的营销客群有较大的响应率及购买金额提升。
	1.5 创新性描述	项目主要运用 “联邦学习”技术，旨在建立基于跨企业数据集的机器学习模型。在模型训练的过程中，模型相关的参数信息能够在各方之间以加密形式交换，但描述用户行为及特征的原始数据不参与交换。建好的模型仅在参与各方本地为各自的业务目标服务。 现有的机器学习模式多为单侧机器学习模式，仅利用了企业自身掌握的数据进行模型训练，如何利用外部数据提升模型的精度在金融行业一直是热门的研究领域。目前业界存在两种常见的模式，其一是通过购买外部数据使用权限作为特征数据进行模型训练的“外部数据”模式。其二是一种被称为“联合建模”的合作式建模方式，常见于个人信贷领域。由信贷业务方将放贷结果数据作为目标变量提供给第三方特征数据持有者进行建模，并将模型结果返回给信贷业务方。 以上两种传统模式中，“外部数据”的提供方在市场上极为有限，对模型算法无法形成大规模的数据维度扩充。“联合建模”模式无法同时利用自有数据及外部数据，需要进行二次建模，降低了模型的准确性。同时由于涉及到目标变量的传输，仍存在客户数据泄露的风险。相比之下，联邦学习技术可以在建模过程中克服数据泄露以及自有数

		据、外部数据不能同时使用的困难。具体来讲： 1、同时使用各参与方的所有数据，能够提供更高的模型精度。参与联邦学习的各数据提供方均可从机制中获益。 2、底层数据不参与传输，数据泄露风险更低。 3、参与各方可在数据源方以及数据应用方之间转换，均可获得业务价值。合作过程中，各方是对等的，不存在一方主导另外一方。以本次合作为例，建信基金作为业务需求提出方，提供模型训练目标变量以及建模特征，合作方银行提供建模特征。合作方银行可以在未来其他项目中作为业务需求提出方，利用搭建好的联邦学习框架，由建信基金作为建模特征提供方提升其模型的精准度。
	1.6 应用价值描述	宏观方面： 机器学习算法在精准营销业务方面已经在互联网、零售行业以及金融行业有了广泛的应用，在提升目标客群的筛选精准度、降低营销成本方面发挥了重要价值。随着数据信息的积累，机器学习算法在基金公司的数字化经营能力建设中也开始扮演重要角色。 随着市场竞争的加剧，基金管理公司对机器学习算法的精度有了更加严格的要求。不同企业、团体存储的海量数据信息受隐私保护和监管的要求，难以联通形成了大量的数据孤岛，从而成为了模型精度提升的新瓶颈。“联邦学习”技术在保护用户隐私以及底层数据安全的前提下，同时利用多方数据特征进行模型训练，为人工智能打破数据屏障、提升营销客服效率方面提供了新的思路。 微观方面： 联邦学习技术分为两类：横向联邦学习技术适用于参与者的数据特征重叠较多，而样本 ID 重叠较少的情况；纵向联邦学习技术适用于参与者的样本 ID 重叠较多，而数据特征重叠较少的情况。合作方银行拥有海量客户群体和丰富的数据维度，对客户在银行视角刻画的刻画非常全面；建信基金深耕基金投资领域，积累了数千万客户全面且深刻的基金相关数据，能够全面描述客户的投资特征。双方数据存在维度互补。同时，双方存在大量代销渠道共有客户。大量的共有客户以及数据维度的互补，故而适用纵向联邦学习技术。 本次试点项目打通了纵向联邦学习技术从理论到实践的路径。试点项目中参与双方为基金公司与一家银行类基金销售机构，此类业务模式可以推广至基金公司与其他具有一定体量的基金销售机构，提升基金销售机构对客户基金产品需求的精准匹配能力，从而为基金投资者提供更优质的服务。 另一种未在本次试点项目中尝试的横向联邦学习技术因其能够为模型训练扩展训练样本量，适用于多家基金公司之间进行合作的场

		景。由于证监会对各家基金公司的系统和数据有大量统一的要求和规定，使得不同基金公司在用户行为数据的标准上有很多共同之处，为未来多家基金公司之间就横向联邦学习技术进行合作尝试提供了良好的先决条件，从而为各基金公司，尤其其中的中小基金的客户服务能力的提升提供了新的尝试方向。
	1.7 试点目的描述	对机器学习前沿的纵向“联邦学习”技术在基金管理公司的精准营销场景下从理论到实践的路径进行探索。对隐私保护前提下不同企业实体进行数据融合应用的可行性，以及“联邦学习”框架下机器学习模型效果的提升进行了可行性测试，为未来在越来越严格的数据监管下开展稳定、持续的外部数据合作建立了模板。同时为行业内基金公司与基金销售机构的数据融合合作打下了基础。
	1.8 牵头申报单位	建信基金管理有限责任公司，基金管理公司
	1.9 联合申报单位	无
二、项目基本信息	2.1 功能服务	本项目将联邦学习的技术应用于智能营销场景，通过历史大数据预测用户未来行为，精准筛选出营销活动的目标客群，从而降低营销成本。项目参与双方的数据分析师及建模师通过联邦学习软件进行共有用户的加密对齐，模型中间计算结果数据的加密传输并对模型进行训练调优，针对市场营销团队的业务需求提供智能营销模型解决方案。 联邦学习技术可以在训练过程中同时利用参与双方的数据特征，对模型精度有进一步的提升。相比之下，传统的单侧机器学习模式不涉及企业间数据传输；联合建模模式则仅涉及共有用户的目标变量以及模型评分结果的传输，数据传输量远远小于联邦学习。两类传统模式均无法在模型训练中同时利用不同企业、团体的特征数据，模型的精准度会受到限制。
	2.2 技术应用	本试点项目使用了机器学习分支中隐私计算领域中的“联邦学习”的新兴技术，该技术通过加密算法，实现在不泄露本方客户数据和不违反国家数据保护法律法规的前提下进行跨企业、实体间的建模工作，因其在模型训练中同时利用了不同企业间的特征数据，能够实现模型精度的提升。相比于随机选取客群或根据业务经验选取目标客群两种传统的营销策略，通过使用模型对客户的行为进行预测评分，更精准的定位目标客群，从而提升营销活动的效率，降低营销成本。 与传统方式相比，联邦学习技术在应用中增加了“样本加密对齐”与“加密模型训练”两个步骤，保证数据隐私安全。首先，在“样本加密对齐”步骤中，联邦学习机制通过融合 RSA 与哈希技术进行样本对齐，使得参与双方在获取共有客户的同时无法从对方获取非自身客

		户的 ID 信息。在此基础上，参与双方分别在各自本地进行数据清洗、筛选工作，并将数据传输至联邦学习软件进行 “加密模型训练”步骤。在两个步骤中，加密后的中间计算结果均存储于模型训练发起方的服务器中，即有业务需求并提供模型训练目标变量的参与方。在本项目中，建信基金作为业务需求提出方，提供服务器存储加密后的中间计算结果。存储的加密中间结果为经过加密算法和聚合算法处理过的梯度值，从中反推回原始数据的可能性目前仅存在于理论中，且需要大量计算资源及大量计算时间，极大的降低了客户隐私数据的泄露风险。 相比传统模式，联邦学习采用了同态加密以及多项式近似的方式对机器学习模型训练中产生的重要中间结果梯度以及损失函数进行了加密并传输，从而使得模型训练中可以同时利用各参与方的特征数据进行训练。同时，加密与多项式近似算法均已被证明对模型结果几乎没有损失。
	2.3 数据应用	本试点项目使用了来自建信基金数据中台以及合作方银行大数据平台两方个人客户的数据，具体包括： 合作方银行大数据平台个人客户包含了、位置信息、客户金融资产分布、个人客户持有产品信息、个人客户消费行为、客户代发工资信息等类别的数据。 建信基金数据来源于建信基金数据中台。数据中台基于大数据平台建设，大数据平台为建信基金公司级的数据中心，整合了公司内部几乎所有的核心业务系统数据、手工维护数据（无相应业务系统或封闭的业务系统）以及按需整合万得、财汇等资讯数据。大数据平台通过批量的方式采集数据，采用 Hadoop 和 Oracle 混搭的结构，其中 Hadoop 共 8 个数据节点，单节点存储约 7TB，Oracle 数据规模 2TB。 为保护数据安全，建信基金建立了全面有效的数据治理组织架构，制定了《建信基金管理有限责任公司数据管理制度》，主要内容包括数据维护、数据备份、备份数据保管和数据加密。由于联邦学习算法在模型训练过程中涉及加密的中间数据传输，建信基金作为业务需求提出方，提供服务器存储加密后的中间计算结果。同时在测试期间确保客户隐私数据以及中间计算数据在传输过程中不使用明码且无法反解。
	2.4 服务对象与渠道	建信基金与合作银行双方的数据分析师以及建模师利用联邦学习平台为建信基金市场营销团队提供算法模型服务，同时提供营销客群清单筛选及短信推送服务。最终的营销服务通过短信的方式对通过持有建信基金“速盈”产品的银行代销个人客户开展营销活动。由于产品的货币属性，短信营销活动符合营销适当性。

	2.5 业务规模	上线后目前的用户为市场营销部，预期在未来扩展到建信基金各分公司以及营销中心。营销服务对象为通过银行代销渠道持有 500 万“速盈”产品的个人客户。
	2.6 预期效果	在此项目中，我们提出了五各联邦学习建模场景，分别从“高净值”、“临界”、“长尾”、“休眠”、“流失”五个方面进行建模，以提升客户的贡献和价值。模型完成后，在单边模型效果已经较好的前提下，联邦学习模型效果全面优于单边模型，AUC 平均提升 10.4%，最高提升 41.2%，KS 平均提升 58.1%，最高提升达到 248%。模型投入营销后期间内，共有 4 万余位客户购买营销目标基金产品，申购金额达超过 10 亿元，营销响应率 5.33%，远超行业平均水平。
	2.7 已获专利、认证或奖项	在 2020 年 12 月，由工信部下辖中国信息通信研究院、中国通信标准化协会大数据技术标准推进委员会联合组织开展的“2020 大数据星河案例评比”中，本项目荣获隐私计算优秀案例。
三、依法合规原则评估	3.1 涉及的业务场景是否由持牌机构提供	是
	3.2 是否违反现行法律法规和监管规定	否
	3.3 分析及结论： 1、由于联邦学习算法在模型训练过程中涉及加密的中间数据传输，公司履行《证券基金经营机构信息技术管理办法》，对本项目中使用的联邦学习框架的底层加密算法以及市场使用情况进行了深层调研，确保了客户隐私数据在模型训练过程中不泄露。 2、公司履行《证券基金经营机构信息技术管理办法》，明确了信息技术管理制度和操作流程，根据本项目的探索性质进行了相应的服务器部署、算力投入以及数据传输专线带宽设置，保障了与业务活动规模及复杂程度相适应的信息技术投入水平。 3、在本项目进展过程中，建信基金与合作银行在项目中任务发起两端的角色对平台的使用权限制定了严格界限，同时角色以及权限均是在审批流程结束后才进行授权。综上所述，本试点项目符合依法合规原则。	
四、有序创新原则评估	4.1 是否侧重于大数据、云计算、人工智能、区块链等新一代信息技术对资本市场各类业务的科技赋能	是
	4.2 是否以服务实体经济、提升市场效能、强化合规风控、增强监管能力、保障金融安全为应用导向	是
	4.3 是否有助于稳妥推动新一代信息技术在资本年市场的落地实施，促进资本市场的数字化发展	是
	4.4 分析及结论： 1、建信基金与合作方商业银行双方均从大数据平台技术以及人才引进起始布局，多年来各自完成了对客户全方位的数据积累以及高效的数据服务能力，为数据分析以及	

<table>
<tr><td></td><td colspan="2">模型服务提供了数据基础。
2、双方在传统的单侧机器学习模式都有较多的积累，数据分析以及算法开发平台均有成熟的部署，为联邦学习技术的尝试打下了算法能力基础。
3、双方长期以来一直开展大数据合作，积累了丰富的数据应用合作经验，为此次的联邦合作提供了合作基础。
综上所述，本试点项目符合有序创新原则。</td></tr>
<tr><td rowspan="4">五、风险可控原则评估</td><td>5.1 是否已有效识别相关业务合规、系统安全、数据安全风险</td><td>是</td></tr>
<tr><td>5.2 是否不存在重大风险隐患或已充分做好相应风险防范和补偿安排</td><td>是</td></tr>
<tr><td>5.3 是否不存在发生系统性风险的隐患</td><td>是</td></tr>
<tr><td colspan="2">5.4 分析及结论：
1、由于联邦学习算法在模型训练过程中涉及加密的中间数据传输，建信基金作为业务需求提出方，提供服务器存储加密后的中间计算结果。同时建信基金存储的加密中间结果为经过加密算法和聚合算法处理过的梯度值，从中反推回原始数据的可能性目前仅存在于理论中，且需要大量计算资源及大量计算时间，极大的降低了客户隐私数据的泄露风险，实现了“数据可用不可见”。
2、双方的业务系统间采用了逻辑隔离的方式相互隔离，仅通过一对一专线进行数据传输，最大化保护用户隐私及数据安全。
3、模型结果的应用与日常生产系统解耦，按营销活动分批次应用，不影响日常系统运行。
4、模型成果在应用于营销活动前，需经过负责测试、数据分析以及营销活动策划的相关人员进行评估把关，确保模型的合理使用并控制模型使用风险。
综上所述，本试点项目符合风险可控原则。</td></tr>
<tr><td>六、业务风险控制机制</td><td colspan="2">外部市场环境的不确定性以及外部竞争的加剧，可能导致营销活动受阻、失败或达不到预期营销的目标等风险。同时联邦学习是根据历史数据建立的模型，模型上线时进行了历史规律在未来仍然适用的假设，从而存在外部环境变化导致模型准确性下降的风险。针对上述风险，我们针对每一次营销活动进行效果追踪并进行归因分析，根据营销活动的追踪分析结果对模型进行优化。</td></tr>
<tr><td>七、技术安全保障机制</td><td colspan="2">试点项目中可能存在计算服务器以及双方数据传输线路故障风险，即服务器或专线传输意外中断导致项目开发延时的风险。针对上述风险，我们对服务器以及传输专线流量日志进行实时监督，将服务器及专线发生的异常情形通过短信以及邮件形式发送至相关技术人员，以便迅速发现查明此技术风险。同时，在服务器运行或专线传输出现异常时，立即切换到备用服务器或传输专线，降低负面影响。</td></tr>
<tr><td rowspan="2">八、投资者保护</td><td>8.1 客户投诉渠道</td><td>本项目不涉及。</td></tr>
<tr><td>8.2 投诉处理机</td><td>本项目不涉及。</td></tr>
</table>

	制		
	8.3 风险补偿机制	本项目不涉及。	
	8.4 项目退出机制	当模型准确性下降时，对联邦学习模型进行下线处理，并对营销活动进行暂停或根据业务规则进行。同时进行线下的模型优化及重新开发。待模型效果提升至可接受范围内后，对模型进行重新上线。	
九、申报单位基本信息	9.1 牵头申报单位	9.1.1 单位名称	建信基金管理有限责任公司
		9.1.2 单位类型	基金管理公司
		9.1.3 统一社会信用代码	91110000717859226P
		9.1.4 注册地址（办公地址）	北京市西城区金融大街 7 号英蓝国际金融中心 16 层
		9.1.5 持有金融牌照情况	建信基金具有： 1、中国证券监督管理委员会颁发的经营证券期货业务许可证（流水号 000000028825），证券期货业务范围包括公开募集证券投资基金管理、基金销售、特定客户资产管理 2、北京市工商行政管理局颁发的营业执照（编号 104446273），经营业务包括基金募集、基金销售、资产管理和中国证监会许可的其他业务。
		9.1.6 试点项目涉及的业务牌照	本项目涉及“基金销售”业务资格
		9.1.7 工作分工	建信基金在本项目中负责基金侧的架构部署、提出业务需求、客户画像分析、数据准备以及模型评估方案制定。 合作方商业银行在本项目中负责银行侧的架构部署、业务需求确认、目标变量定义、数据准备以及营销短信分发。
		9.1.8 单位简介	建信基金管理公司成立于 2005 年 9 月，由建设银行控股，其他股东包括美国信安金融集团和中国华电集团资本控股公司，是我国首批由商业银行试点设立的基金

			管理公司之一。 历经十五年发展，建信基金已成长为在基金行业内具有重要影响的大型基金管理公司。截至2021年3月底，公司累计为超过4461万客户提供全方位理财服务，资产管理规模1.42万亿元。十五年来，公司旗下基金累计实现分红948.91亿元，为客户创造了良好回报。公司严格依法纳税，纳税额逐年攀升，公司和员工累计纳税超52亿元，并被北京市评为“纳税信用A级企业”。 公司在全国各地设立了5家分公司和6个营销中心，一家境内控股子公司--建信资本管理公司，一家境外全资子公司--建信资产管理（香港）有限公司。公司高度重视人才队伍建设，吸引和培养高素质员工队伍，现有员工622人，其中73%以上具有硕士及以上学历，25%以上具有海外留学或工作经历。
	9.2联合申报单位1	9.2.1单位名称	无
		9.2.2单位类型	无
		9.2.3统一社会信用代码	无
		9.2.4注册地址（办公地址）	无
		9.2.5持有金融牌照情况	无
		9.2.6试点项目涉及的业务牌照	无
		9.2.7工作分工	无
		9.2.8单位简介	无
十、其他补充事项	无		

十一、其他申报材料清单	材料名称	出具单位（部门）	时间（时效）
十二、牵头申报单位承诺	本单位郑重承诺： 1. 本单位在申报资本市场金融科技创新试点（北京）项目过程中，所提供的一切申报材料信息真实、准确和完整，本单位承诺承担与此相应的法律责任。 2. 申报项目符合依法合规、有序创新、风险可控的申报原则。 3. 申报项目不存在违法法律和行政法规情况，不包含国家秘密信息。 4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。 单位（公章） 法定代表人（签字）： 2021 年 11 月 10 日		

附页：联合申报单位承诺

项目名称	
联合申报单位承诺1	本单位郑重承诺： 1. 本单位在申报资本市场金融科技创新试点（北京）项目过程中，所提供的一切申报材料信息真实、准确和完整，本单位承诺承担与此相应的法律责任。 2. 申报项目符合依法合规、有序创新、风险可控的申报原则。 3. 申报项目不存在违法法律和行政法规情况，不包含国家秘密信息。 4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。 单位（公章） 法定代表人（签字）： 年　　月　　日