

Classification		Content
Course scope & objective	Course objective & content	<ul style="list-style-type: none"> - Learning automatic classification model to forecast customers' satisfaction with financial products for a month in a bank - Learning automatic classification algorithm search and application including various types of data - Learning system evaluation method (e.g., cross validation) and evaluation index (e.g., MAP, Precision@K, Recall@K, nDCG, Kendal's tau) - Recommending financial products based on forecast model
	Hands-on context	<p>Learners study automatic classification algorithm and creative improvement:</p> <ul style="list-style-type: none"> - To develop an algorithm recommended by bank and online shopping mall - To select proper algorithm according to the types encoded in stats and data mining solution, such as SPSS and SAS - To search the optimal algorithm after evaluating various algorithms
Idea of hands-on problems		<ul style="list-style-type: none"> - Which factor has a high correlation with a financial product in a month: such as age, employment, residential district, deposit in account, mortgage, credit card etc.? - How can we finally forecast the financial product based on the correlation above? - How can we do machine perception learning if we have a big data?
Learner's role in hands-on context		a Software College student who has a cousin working in a bank
Scenario of hands-on problems		To forecast a future financial product from the data Santandar bank customers
		<p>Mr. B, a person in charge of managing customer data in the bank, has been directed to forecast a future financial product. Although the bank has a license of SPSS and SAS, the big data for sex, age, and amount of deposit were not identified; moreover, there were no algorithm to forecast them correctly. Therefore, Mr. B wants to seek advice from his cousin who is studying software at Hanyang University.</p>

Item		Description
Problem Formulation	Given data and our goal	<ul style="list-style-type: none"> We are given 6 files of customer data of a bank; each file is customer data recorded at every 28th day from 2015-01-28 to 2015-06-28. In each data file (in csv format), each row include customer's attributes (e.g., gender, age, etc.) and the financial product he/she is using at the day when data is recorded. There is four financial product (e.g., saving_account, fund, mortgage, credit_card) and the field of each item is 0 or 1, which represents if he/she is using the product. Our goal is to for each product, predict the chance that a customer newly purchases it at 2015-07-28 based on the given data and select the top-100 customers who is expected to NEWLY buy it. In the test data file 2015-07-28.csv, those four columns on financial products are empty. <p>Given files</p> <ul style="list-style-type: none"> training data set: 2015-01-28.csv ~ 2015-06-28.csv test data set: 2015-07-28.csv
		<p>1. result1.txt ~ result4.txt: each file includes 50 lines</p> <p>result1.txt => about saving_account result2.txt => about fund result3.txt => about mortgage result4.txt => about credit_card</p> <p>output in each line: <the customer id></p> <p>IMPORTANT: list 50 customers in each file in the order of decreasing probability that they buy the product</p> <p>2. source files</p>
		<p>Mean Average Precision: the mean of AP@50 of 4 products where</p> $AP@50 = \frac{1}{\min(m, 50)} \sum_{k=1}^{\min(m, 50)} P(k)R(k)$ <ul style="list-style-type: none"> m: the number of new buyer of each product P(k): Precision@k - the ratio of new buyer out of k R(k): indicator function returning 0 or 1 according to whether or not the customer in the k-th line newly purchased the product

Item	Description
Attributes	Column Name Description id Customer code employee Employee index: A active, B ex employed, F filial, N not employee, P pasive gender Customer's sex age Age since_first_date Months since the customer became a contract holder cust_type Customer type at the beginning of the month, 1 (First/Primary customer), 2 (co-owner),P (Potential), 3 (former primary), 4(former co-owner) res_index Residence index (S (Yes) or N (No) if the residence country is the same than the bank country) for_idx Foreigner index (S (Yes) or N (No) if the customer's birth country is different to the bank country) sp_index Spouse index. 1 if the customer is a spouse of an employee channel channel used by the customer to join res_code Province code (customer's address) active Activity index (1, active customer; 0, inactive customer) income Gross income of the household cate Category: 01 – VIP, 02 – Individuals 03 – college graduated saving_account Saving Account fund Funds mortgage Mortgage credit_card Credit Card