METHODS AND SYSTEMS FOR AUTOMATICALLY GENERATING HIGH QUALITY ADVERSE ACTION NOTIFICATIONS

Applicant: ZESTFINANCE, INC., Los Angeles, CA (US)

Inventors: John W.L. Merrill, Redmond, WA (US); Shawn M. Budde, Evanston, IL (US); John B. Candido, III, Burbank, CA (US); Lingyun Gu, Thousand Oaks, CA (US); Farshad Kheiri, Woodland Hills, CA (US); James P. McGuire, Long Beach, CA (US); Douglas C. Merrill, Los Angeles, CA (US); Manoj Pinnamaneni, Los Angeles, CA (US); Marick Sinay, Hawthorne, CA (US)

Assignee: ZESTFINANCE, INC., Los Angeles, CA (US)

Filed: Jan. 31, 2014

Abstract

This invention relates generally to the personal finance and banking field, and more particularly to the field of lending and credit notification methods and systems. Preferred embodiments of the present invention provide systems and methods for automatically generating high quality adverse action notifications based on identifying variations between a declined borrower’s profile and that of approved applicants, both with simple and sophisticated credit scoring systems, using specific algorithms.
FIGURE 5a

301
ID

302
Score

303
Exchange list

(a) Single exemplar with associated exchange score
FIGURE 5b

Exchange list 1  Exchange list 2  Exchange list 3

Exchange list 1  Exchange list 1  Exchange list 3

Original selection 310

Revised ranking by scoring 311
METHODS AND SYSTEMS FOR AUTOMATICALLY GENERATING HIGH QUALITY ADVERSE ACTION NOTIFICATIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application, No. 61/454,496, entitled “METHODS AND SYSTEMS FOR AUTOMATICALLY GENERATING HIGH QUALITY ADVERSE ACTION NOTIFICATIONS” filed Jan. 31, 2013, which application is hereby incorporated in its entirety by reference. This application also relates to U.S. application Ser. No. 13/454,970, entitled “SYSTEM AND METHOD FOR PROVIDING CREDIT TO UNDERSERVED BORROWERS” filed Apr. 24, 2012, which application is also hereby incorporated in its entirety by reference.

TECHNICAL FIELD

[0002] This invention relates generally to the personal finance and banking field, and more particularly to the field of lending and credit notification methods and systems.

BACKGROUND AND SUMMARY

[0003] People use credit daily for purchases large and small. Lenders, such as banks and credit card companies, use “credit scores” to evaluate the potential risk posed by lending money. These measures purport to determine the likelihood that a person will pay his or her debts.

[0004] But credit scoring systems are a recent innovation. In the 1950’s and 1960’s, credit decisions were made by bank credit officials, who knew the applicant. Since applicants usually lived in the same town, credit officers would make subjective decisions based on their knowledge of the applicant. Then, in the 1970’s, the advent of “FICO scores” made credit far more available, effectively removing the credit officer from the process. As the idea of using statistical computations to measure risk caught on, lenders began using a variety of other credit scoring methods and models. Examples include simple systems (such as those offered by Experian, Equifax, TransUnion) and advanced models with many traditional and non-traditional variables, and many thousands of meta-variables, such as those described by the Applicant in U.S. application Ser. No. 13/454,970 and its related continuations.

[0005] As a result of decreased physical interaction with lenders and increased computational complexity of calculating “creditworthiness,” consumers’ visibility into what “drives” their credit scores diminished.

[0006] In response, laws such as the Consumer Credit Protection Act and Federal Equal Credit Opportunity Act, were designed, in part, to ensure that borrowers are not denied loans for discriminatory purposes. They were also designed, in part, to educate consumers as to the reasons why they were denied loans.

[0007] Federal and some state laws require lenders send “adverse action letters” to denied applicants explaining the reasons for those decisions. Theoretically, the letter should also allow an opportunity for the borrower to determine (1) if there is an error in his or her records which has led to the denial of credit; as well as (2) provide the borrower with information that helps identify which part(s) of his/her credit history is problematic.

[0008] But lenders engage in only minimal compliance, often providing generic reasons for their decisions. Unfortunately, these adverse action letters do very little to actually help consumers verify their credit history and/or determine which actions could increase their creditworthiness (at least, insofar as these mathematical models are concerned). This outcome is not beneficial from a lending policy standpoint as customers aren’t getting what they need.

[0009] However, generating adverse actions letters that provide useful consumer feedback is not a straightforward task. Describing market factors as a reason for denials, in and of itself, is a complex job. But when compounded with the complexity of the newer mathematical credit scoring models, the job of effectively communicating the reasons is far more challenging. Indeed, pinpointing one or more variables—jointly or severally—that correlate to increased credit scores involves a complex analysis that few lenders, if any, would bother to perform, much less communicate to their consumers.

[0010] Accordingly, improved systems and methods for generating high quality adverse action letters would be desirable.

SUMMARY OF THE INVENTION

[0011] To improve upon existing systems, preferred embodiments of the present invention provide a system and method for automatically generating high quality adverse action notifications. One preferred method for automatically generating high quality adverse action notifications can include entering and/or importing a borrower dataset and a lender’s credit criteria at a first computer (borrower data and lender criteria); processing the dataset variables and/or sets of variables in the lender’s algorithms to identify which variables, when changed, result in an increased credit score (field selection); ranking individual variables and/or sets of variables in the borrower dataset to yield the greatest differences in a credit score (field ranking); and generating a report showing which variables and/or sets of variables, when changed, result in an acceptable credit score (reason test generation).

As described below, the preferred method can further include formatting the reason set generation into an adverse action letter that is understandable and usable by the consumer (adverse action letter generation). Other variations, features, and aspects of the system and method of the preferred embodiment are described in detail below with reference to the appended drawings.

[0012] The present invention could be used independently (by simply generating adverse action letters) or in the alternative, the present invention could also be interfaced with, and used in conjunction with, a system and method for providing credit to borrowers. An example of such systems and methods is described in U.S. patent application Ser. No. 13/454,970, entitled “System and Method for Providing Credit to Underserved Borrowers,” to Douglas Merrill et al, which is hereby incorporated by reference in its entirety (“Merrill Application”).

[0013] Other systems, methods, features and advantages of the invention will be or will become apparent to one with skill in the art upon examination of the following figures and detailed description. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the invention, and be protected by the accompanying claims.
BRIEF DESCRIPTION OF THE FIGURES

[0014] In order to better appreciate how the above-recited and other advantages and objects of the inventions are obtained, a more particular description of the embodiments briefly described above will be rendered by reference to specific embodiments thereof, which are illustrated in the accompanying drawings. It should be noted that the components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention. Moreover, in the figures, like reference numerals designate corresponding parts throughout the different views. However, like parts do not always have like reference numerals. Moreover, all illustrations are intended to convey concepts, where relative sizes, shapes and other detailed attributes may be illustrated schematically rather than literally or precisely.

[0015] FIG. 1 is a diagram of a system for automatically generating high quality adverse action notifications in accordance with a preferred embodiment of the present invention.

[0016] FIG. 2 depicts an overall flowchart illustrating an exemplary embodiment of a method by which high quality adverse action notifications are automatically generated.

[0017] FIG. 3 depicts a flowchart illustrating an exemplary embodiment of a method for important field selection.

[0018] FIG. 4 depicts a flowchart illustrating an exemplary embodiment of a method for finding the path to adequacy.

[0019] FIG. 5a depicts a flowchart illustrating an alternative exemplary embodiment short titled “swapping codes” as contained in the method for determining the path to adequacy.

[0020] FIG. 5b depicts a flowchart illustrating an alternative exemplary embodiment short titled “selection by scoring” as contained in the method for determining the path to adequacy.

[0021] FIG. 5c depicts a flowchart illustrating an alternative exemplary embodiment short titled “mutation” as contained in the method for determining the path to adequacy.

[0022] FIG. 5d depicts a flowchart illustrating an alternative exemplary embodiment short titled “cross-over” as contained in the method for determining the path to adequacy.

DEFINITIONS

[0023] The following definitions are not intended to alter the plain and ordinary meaning of the terms below but are instead intended to aid the reader in explaining the inventive concepts below:

[0024] As used herein, the term “BORROWER DEVICE” shall generally refer to a desktop computer, laptop computer, notebook computer, tablet computer, mobile device such as a smartphone or personal digital assistant, smart TV, gaming console, streaming video player, or any other, suitable networking device having a web browser or stand-alone application configured to interface with and/or receive any or all data to/from the CENTRAL COMPUTER, USER DEVICE, and/or one or more components of the preferred system 10.

[0025] As used herein, the term “USER DEVICE” shall generally refer to a desktop computer, laptop computer, notebook computer, tablet computer, mobile device such as a smartphone or personal digital assistant, smart TV, gaming console, streaming video player, or any other, suitable networking device having a web browser or stand-alone application configured to interface with and/or receive any or all data to/from the CENTRAL COMPUTER, BORROWER DEVICE, and/or one or more components of the preferred system 10.

[0026] As used herein, the term “CENTRAL COMPUTER” shall generally refer to one or more sub-components or machines configured for receiving, manipulating, configuring, analyzing, synthesizing, communicating, and/or processing data associated with the borrower and lender. Any of the foregoing subcomponents or machines can optionally be integrated into a single operating unit, or distributed throughout multiple hardware entities through networked or cloud-based resources. Moreover, the central computer may be configured to interface with and/or receive any or all data to/from the USER DEVICE, BORROWER DEVICE, and/or one or more components of the preferred system 10 as shown in FIG. 1. The CENTRAL COMPUTER may also be the same device described in more detail in the Merrill Application, incorporated by reference in its entirety.

[0027] As used herein, the term “BORROWER’S DATA” shall generally refer to the borrower’s data in his or her application for lending as entered into by the borrower, or on the borrower’s behalf, in the BORROWER DEVICE, USER DEVICE, or CENTRAL COMPUTER. By way of example, this data may include traditional credit-related information such as the borrower’s social security number, driver’s license number, date of birth, or other information requested by a lender. This data may also include proprietary information acquired by payment of a fee through privately or governmentally owned data stores (including without limitation, through feeds, databases, or files containing data). Alternatively, this data may include public information available on the internet, for free or at a nominal cost, through one or more search strings, automated crawls, or scrapes using any suitable searching, crawling, or scraping process, program, or protocol. Moreover, borrower data could include information related to a borrower profile and/or any blogs, posts, tweets, links, friends, likes, connections, followers, followings, pins (collectively a borrower’s social graph) on a social network. The list of foregoing examples is not exhaustive.

[0028] As used herein, the term “LENDER CRITERIA” shall generally refer to the criteria by which a lender decides to accept or reject an application for credit as periodically set in the USER DEVICE or CENTRAL COMPUTER. By way of example, these criteria may include accept or reject criteria based on individual data points in the BORROWER’S DATA (such as length of current residence—6 months), or based on complex mathematical models that determine the creditworthiness of a borrower.

[0029] As used herein, the term “NETWORK” shall generally refer to any suitable combination of the global Internet, a wide area network (WAN), a local area network (LAN), and/or a near field network, as well as any suitable networking software, firmware, hardware, routers, modems, cables, transceivers, antennas, and the like. Some or all of the components of the preferred system 10 can access the network through wired or wireless means, and using any suitable communication protocol/s, layers, addresses, types of media, application programming interface/s, and/or supporting communications hardware, firmware, and/or software.

[0030] As used herein and in the claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise.

[0031] Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art.
DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0032] The following description of the preferred embodiments of the invention is not intended to limit the invention to these preferred embodiments, but rather to enable any person skilled in the art to make and use this invention. Although any methods, materials, and devices similar or equivalent to those described herein can be used in the practice or testing of embodiments, the preferred methods, materials, and devices are now described.

[0033] The present invention relates to improved methods and systems for automatically generating high quality adverse action notifications, which includes notifications for individuals, and other types of entities including, but not limited to, corporations, companies, small businesses, and trusts, and any other recognized financial entity.

[0034] Prior to describing the preferred system and method, a frame of reference is in order: As many consumers know, a person’s credit history is made up of a number of variables, such as the amount of debt the person is presently carrying, their income stability, their repayment history on past debt (lateness or failure to pay), and the length of their credit history. However, consumers do not often appreciate that modern credit scoring systems have significantly increased in sophistication, now containing many variables and meta-variables as well.

[0035] But credit applications are typically underwritten in a fairly stereotyped fashion. First, information is gathered from an applicant, generating an application. This application is supplemented with other data relevant to the application, including publicly available data (e.g. “Does the applicant have an active lien filed against him/her?”) or proprietary data (e.g. “How many credit applications has this person filed in the last 90 days?”) The resulting record is provided to a “scoring system,” which assigns a numerical score to the application. This score is compared to a threshold, and the applicant receives a loan if the score exceeds that threshold.

[0036] Given this process, the problem of determining the reasons for a declined application appears straightforward: simply go through the list of variables (also called “signals”) used in the scoring process, and find the ones which could make the score change towards the threshold. This appearance can be deceptive.

[0037] Modern scoring functions almost never produce monotone base signals. Even scoring systems that use monotone functions such as logistic regression to produce their final outputs use intermediate meta-variables that transform primary signals into distilled signals that are no longer monotone in the values of those primary signals. As a result, knowing that a small change in a given signal locally increases the application’s score for a given application gives no assurance that a large change in the same signal will necessarily still increase the application’s score. In addition, since a change in a single variable might not have the same effect as a change in several variables together, it would not necessarily be enough to look at one-variable perturbations when searching for “better results.”

[0038] Because detection of the underlying variables that drive an applicant’s score has become exceedingly complex, in turn, intelligibly reporting the right reasons for deficiency in a person’s score requires not only sophisticated analysis, but also a translation into understandable language. Indeed, it is because of this complexity that few lenders, if any, would bother to perform the analysis, much less communicate to their consumers. The preferred embodiment of the present invention solves this problem through an automated method and process for performing the analysis and intelligibly communicating the results, as further described below:

System

[0039] As shown in FIG. 1, a preferred operating environment for automatically generating high quality adverse action notifications in accordance with a preferred embodiment can generally include data sources (the borrower’s application 13, and the lender’s credit model 15), a USER DEVICE 30, a CENTRAL COMPUTER 20, a NETWORK 40, and one or more communication devices from which the borrower is issued an adverse action letter, including a BORROWER DEVICE 12, Email Server 30, and/or a Print Server 40.

The preferred system 10 can include at least: data sources (the borrower’s application 13, and the lender’s credit model 15), and a computer to analyze and process the data sources (CENTRAL COMPUTER 20 and/or a USER DEVICE 30), which function to generate high quality adverse action notifications.

To be clear, the borrower’s application 13 should include one or more variables in the BORROWER DATA, and the lender’s credit model 15 should include one or more algorithms from the LENDER’S CRITERIA. In particular, the preferred system 10 functions to help borrowers determine the accuracy of his/her credit file as well as provide information to improve his/her creditworthiness, by accessing, evaluating, measuring, quantifying, and utilizing a the novel and unique methodology described below.

[0040] More specifically, this invention relates to the preferred methodology for automatically generating high quality adverse action notifications that takes place within the CENTRAL COMPUTER 20 and/or a USER DEVICE 30, after gathering and/or downloading the BORROWER’S DATA 13 and the LENDER CRITERIA 15.

Method

[0041] FIG. 2 provides a flowchart illustrating one preferred method for automatically generating high quality adverse action notifications which involves the following steps: (a) gathering the BORROWER DATA 100; (b) important field selection 200 (to compare BORROWER DATA against the LENDER CRITERIA 600); (c) field ranking 300; (d) reason text generation 400; and (e) generating adverse action letters 500.

[0042] In the first step, all data from the borrower’s failed application (BORROWER DATA 100) is temporarily gathered for collection by a computer (such as the CENTRAL COMPUTER 20 in FIG. 1). For example, the BORROWER DATA 100 may include classic financial data such as the borrower’s current salary, length of most recent employment, and the number of bankruptcies. Additionally, the BORROWER DATA 100 may include other unique aspects of the borrower, such as the number of organizations the borrower has been or is currently involved with, the number of friends the borrower has, or other non-traditional aspects of the borrower’s identity and history such those identified in the Merrill Application. Subsets of BORROWER DATA 100 are used to determine the borrower’s credit score.

[0043] For illustrative purposes, fictitious BORROWER DATA 100 for Mr. “A” (a creditworthy applicant), Mr. “B” (a declined applicant), the average approved applicant, and the perfect applicant are shown below:
Referring back to FIG. 2, the second step is important field selection 200. Important field selection is the creation of a list of BORROWER DATA variables whose values either reduce or increase the application’s credit score by sufficiently perceptible amounts when those variables are changed, and processed through the LENDER CRITERIA 600.

As shown in FIG. 3, important field selection 200 may be accomplished by determining the shortest path between the borrower’s credit application and the “perfect application” (shortest path 210). Alternatively, important field selection 200 may be accomplished by finding the most important changes between the borrower’s application and an “adequate application” that is approved for funding (path to adequacy 220). Both methods are discussed below:

Starting with the first method, the shortest path 210: as its name implies, the shortest path 210 is a protocol in which a list of all fields (variables) are identified where there is difference between the BORROWER DATA and the data of a “perfect” applicant. Given that a “perfect” application (one which receives the highest possible score) will always be funded, one way to build an explanation for why a different application was not approved is to find the set of differences between the unfunded application and the perfect application. Thus, as a preliminary step, the preferred method is to record a list of fields on which the two applications differ.

By way of example, we again refer to the fictitious credit applicant, Mr. “B” whose shortest path is as follows:

It is important to remember that BORROWER DATA 100 could include dozens variables or hundreds of thousands of meta-variables. And depending on the sophistication of the Lender’s credit scoring system, some or most of those variables and meta-variables may not be used in determining a borrower’s credit score. The shortest path 210, may not be helpful to the applicant in (1) identifying flaws in his credit profile and (2) determining what actions would be necessary to improve his creditworthiness. Thus, if an applicant takes selective actions in “remedying” portions of his/her credit profile; those changes may not result in a score improvement that would meet the LENDER CRITERIA 600.

In other words, the borrower may not be able to recognize which variables are important, and which are just chaff.

As a result, the preferred method in the shortest path 210 includes an intermediate step that eliminates “low impact” fields (which are later omitted from the reason text generation 400, and in turn, the adverse action letter 500 as shown earlier in FIG. 2).

Contrary to logical conclusions, the preferred method for eliminating “low impact” fields does not directly identify “low impact” fields. Rather, the focus is to find the “signals” that are important. And in order to find the signals that are important, the preferred method is to pick the variables which require the smallest transformation (i.e. the shortest path) from a given application to an application with a perfect score. A singular path may be chosen at random with signals then selected based on their relative impact. Alternatively, if multiple paths are available, then lists of variable are ranked by frequency, if possible.

Path-finding is a well-studied problem in machine learning in either a graph or a continuous domain, and there are many well-studied algorithms for finding optimal or near-optimal paths, including, without limitation: ant colony optimization, swarm-based optimization techniques, steepest and stochastic descent algorithms. In addition, there are many multidimensional optimization algorithms available, which has been a major area of study in computer science since the first computer was built. Other path finding algorithms may be used as well depending on suitability to the data set and/or desired outcome. Thus, depending on the nature of the algorithms in the LENDER CRITERIA 600, these path finding algorithms may be applied singularly, or in a hybrid approach, depending on whether the features of the LENDER CRITERIA 600 and/or BORROWER DATA 100 are continuous and/or discrete.

To illustrate the differences between continuous and discrete LENDER CRITERIA and/or BORROWER DATA 100, additional examples may be helpful. A lender criterion might be discrete (e.g., Does the borrower have a job and a checking account?). Similarly, a borrower signal can also be discrete (e.g., is the borrower employed (yes/no)? Does the borrower have a bank account (yes/no)?). Conversely, a lender criterion can be continuous (weight the application negatively according to the average amount of ethanol consumed by the applicant each week). And the corresponding borrower signal would also be continuous (how many beers have you drunk in the last week? Glasses of wine? Mixed drinks/other distilled liquor products?)

The shortest path 210 may be further “filtered” whereby denials for seemingly spurious fields (such as the
number of friends one has in social media), could be eliminated from the important field selection list. FIG. 4 provides a second perspective in illustrating the preferred method to find the shortest path 210. In order to find the shortest path 210, a comparison of known good application(s) 211 would be made against known bad application(s) 212. From this comparison, a list of identical signals 213 and different signals 214 could be obtained. Thereafter, the incremental changes to the variables/fields that produce different signals 214 would be run against a series of selection tests 215. One test might determine if changes to individual variables, or sets of variables, result in a sufficiently improved credit score. A second test may eliminate those fields, that when changed, does not result in substantial improvement—or any improvement—in the applicant’s credit score. A third test may include a manual filter whereby certain variables/fields are eliminated for administrative purposes.

Referring back to FIG. 3, a second preferred method to important field selection may be achieved by finding the most important changes in a path to an adequate application (path to adequacy).

Unlike the shortest path method 210, which generally returns one path (or depending on the tests employed in determining the “perfect application;” a few paths), the path to adequacy 220 is likely to return numerous paths to fundability.

The preferred method for generating the path to adequacy seeks the shortest paths from a given application to applications that have scores exceeding a specified threshold (where the threshold is no greater than the maximum possible value of the scoring function). The methods for doing so are similar to that found in the shortest path 210, except that instead of comparing the borrower’s profile to a perfect application, it is instead compared to a collection of accepted applicants.

There are a lot more path(s) to adequacy than path(s) to perfection. Unfortunately, the range of possible “acceptable” applications usually is not structured nicely. There are many subtle interactions among different signals in an application. This means that some set of signals is likely to occur. Often, however, there are so many sets of changes that it would be effectively impossible to examine each possibility to reach an acceptable application. Instead, the preferred method of the present invention is to identify a set of changes to the failing application when compared to previously collected approved applicants.

Depending on the sophistication of the LENDER CRITERIA, the number of subsets of the list of exchanged fields grows exponentially in the number of fields. It is impossible to enumerate all of them. Instead, the preferred approach is probabilistic: taking random subsets of the set of exchanged fields, and measuring the resulting score change. In such instances, the preferred method is to use the score changes over all samples. The result turns out to be a rough weighting of the contribution of the individual fields to the final score change.

By way of example, we again refer to the fictitious credit applicant, Mr. “B” whose path to adequacy is as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mr. “B”</th>
<th>Sample Applicant</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loan Requested</td>
<td>$7,500</td>
<td>$1,500</td>
<td>$6,000</td>
</tr>
<tr>
<td>Income</td>
<td>$65K</td>
<td>$48K</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Referring back to FIG. 2, the third step is field ranking 300. The preferred approach for field ranking 300 will depend on whether important field selection 200 is accomplished by way of the shortest path method 210 or the path to adequacy 220.

If the shortest path method 210 is employed, ranking, although possible, is purely academic. Indeed, and if well specified, the truncation of the shortest path effectively creates an “all or nothing” result of a long list of fields. In other words, since all changes dictated by the shortest path are necessary to make the application fundable, there is no need to rank the important field selection 200.

If the path to adequacy 220 is employed, the preferred method would regulate the number of fields by ranking the fields so that higher-ranked fields contribute more to a passing score than lower-ranked ones.

Because more than one shortest path to the threshold is likely to exist, the preferred ranking method would employ a voting strategy. In the preferred method, the computer performs many simultaneous searches for many paths to the specified threshold, and then the computer votes based on the number of paths a given field occurs in. Examples include, but are not limited to: membership in the greatest number of paths, chances that have the greatest impact, or some combination thereof. A complete enumeration of the methods is not possible. However, the preferred method will seek to have a meaningful correlation to signal impact, and avoids verging into an arbitrary ranking or scoring function, where possible. Notwithstanding, arbitrary ranking or scoring functions are an alternative method.

For example, assume that Mr. B’s BORROWER DATA has 26 possible paths to adequacy 220. And within those lists, a “loan amount-to-income” meta-variable appears 21 times, the variable “social security numbers” appears in 16 times, and the variable “number of addresses” appears 4 times. Let’s further assume that one of the protocols in the field ranking 300 says that fields that appear in at least 20% of the total paths (or 5 occurrences) should be reported. In this example, “loan amount-to-income” and “number of social security numbers” would be highest, and in this order.

An alternate method to field ranking 300 is to estimate the “contribution” of each field in each path to the final score difference. As stated above, one method to do so is to take random samples of the fields for any given path and compute the score that arose from just using values in those fields, and take the average difference across all paths containing each field as an importance score (while ranking fields according to their importance).

However, the preferred method for identifying “contributions,” (also known as creating “weighted importance scores”) is either accomplished by using (1) a ranking by scoring methodology, or (2) through a genetic algorithm.

In the instant invention, either electronic method can be used to more efficiently select the most regularly
occurring sets of high-impact changes that could be made within a set of paths (or aggregated portions of paths) that result in credit approval.

[0069] The ranking by scoring method significantly reduces the number of searches for adequate paths (or portions thereof) that would lead to an acceptable credit score. Rather than using a purely random selection of variables, the ranking by scoring groups items into small sets to be evaluated tournament style. Thus, by limiting the number of sets that may be grouped, ranking by scoring effectively ranks a limited, yet decreasingly random population of paths, which is thereafter ranked.

[0070] A simple example may provide a helpful background: As shown in FIG. 5a (single associated exchange score), occurs when the values of one set of deficient variables (ID 301), has their values replaced (exchange list 303) which results in a new, and preferably acceptable, resulting credit score (score 302).

[0071] As shown in FIG. 5b, when multiple fields are deficient, ranking may be made by “ranking by scoring.” In essence, ranking importance scores is accomplished by replacing the values in an initial set of variables (original selection 310) with a second set of values (revised ranking by scoring 311), and then by scoring the possible replacements. This process would likely be given a limited universe (e.g., computer, please select 1,000 random sets of variables), then continue exchanging combinations of variables—tournament style—until the most potent changes are identified and ranked.

[0072] In the alternative to ranking by scoring, the use of a genetic algorithm may be employed. Genetic algorithms are a well-studied area of computational science that seeks to generate useful solutions to optimization and search problems. In the instant invention, a genetic algorithm would seek out the “pieces of the paths” that most frequently, and most effectively, produce an acceptable credit score.

[0073] At its core, a genetic algorithm uses the evolutionary processes of crossover and mutation to randomly assemble new offspring from an existing population of solutions. The parent solutions are then “selected” to generate offspring in proportion to their fitness. The more fit, or better matched, to the achieving a credit worthy score, an individual model is, the more often it will contribute its genetic information to subsequent generations.

[0074] In the present invention, a genetic algorithm would first engage in mutation (randomly identifying sets of variables and changing values within those sets of variables), “cross over” the sets of variables (i.e., find the most effective sets of variables and values to change), and then “select” a population of paths that are more impactful than others. This process would be iteratively repeated and optimized through “generations” of changes within the sets of variables to determine how effectively each set of changes lead to a passing credit score. During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (sets of changes that quantitatively produce greater changes in the credit score) are typically more likely to be selected.

[0075] The number of initially selected sets of variables, and generations which traditionally be limited (e.g., computer please pick 1,000 sets of variable or computer please limit your search to 100 generations) to limit processing times.

[0076] To illustrate the concept graphically, mutation is illustrated in FIG. 5c. In essence, mutation replaces the swap point 320 between one exchange list and another.

[0077] As shown in FIG. 5d, “cross-over” replaces an initial set of variables/values (original selection 310) and with a second set of variables/values (revised ranking by scoring 311) by mutating possible replacements amongst various possibilities.

[0078] Generally, the average fitness will increase, since only the best solutions from the first generation are selected for breeding, along with a small proportion of less fit solutions. These less fit solutions ensure diversity within the population of parent solutions and therefore ensure the genetic diversity of the subsequent generation of children.

[0079] In other words, mutation and/or cross-over operate to produce a number of different candidates, which are then ranked by their scores (highest to lowest), and then resampled with a weight according to each score.

[0080] An example may be helpful to further explain mutation and cross-over. Consider a system with five input signals: age, employment status, bank account status, income, and distance between home and work. The LENDER CRITERIA specifies that only applicants with employment and checking accounts are accepted. Moreover, the LENDER CRITERIA will result in a rejection if the applicant earns less than $40,000 per year and/or the applicant lives more than twenty miles from the applicant’s place of employment. Age is completely ignored. Three hypothetical rejected applicants could have the following data:

<table>
<thead>
<tr>
<th>Applicant</th>
<th>Applicant C</th>
<th>Applicant D</th>
<th>Applicant E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>61</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>Employment</td>
<td>No, but . . .</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Checking account</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Income</td>
<td>$50,000 pension</td>
<td>$35,000</td>
<td>$30,000</td>
</tr>
<tr>
<td>Distance</td>
<td>0</td>
<td>25</td>
<td>27</td>
</tr>
</tbody>
</table>

[0081] In the preferred method, the most “important reasons” for each of the three applicants (C, D, and E) would first look for randomly selected sets of swaps. Each of those swaps would then be scored. By comparing each applicant to funded applications, the preferred method would generate a set of frequencies for each variable (or randomly selected set thereof). Using random substitution of values for each variable (or sets of variables) would take an inordinate period of time. Therefore, the preferred method could exchange values individually, or in blocks. This resulting set produces an ordering: application C needed a job and a checking account, application D would need to live closer to work, and application E would need a checking account and to live closer.

[0082] On a broader note, these ranking protocols fall into two categories: continuous parameters and discrete search space. For continuous parameters, algorithms search parameters such as regression and Lyapunov functional reduction are particularly well suited. However, for discrete search space, other suitable search space algorithms, such as pure random search, simulated annealing, and/other genetic algorithms are better suited.

[0083] To recap and further illustrate a practical example of the first three steps in FIG. 2 (Gathering BORROWER DATA 100, performing important field selection 200 by comparing to the LENDER CRITERIA 600, and Field Ranking 300), let’s take the example of Mr. B and compare (1) his scored
application which did not meet the threshold of fundability to (2) other applications previously scored that were fundable. [0084] First, the preferred method is to gather Mr. B’s BORROWER DATA as well as extract the subset of previous applications with scores fundability threshold. Next, and assuming the important Field selection is accomplished by the path to adequacy 220, the preferred method for important field selection 200 is to create an initial population of exemplars consisting of an index into that subset and a bit vector of the same length as the list of features for the LENDER CRITERIA 600. Each exemplar will be scored by taking Mr. B’s un-awarded loan and replacing the list items where the bit vector is 1 with the values from the indexed element of the subset. Finally, the preferred method is to compute the score of Mr. B’s modified list. This process will be iteratively repeated until an appropriate termination criterion has been reached (e.g., all paths to fundability have been identified or the method-defined maximum number of paths has been identified).

[0085] Thereafter, all of the important field selection 200 entries will be field ranked 300. In this step, a new set of exemplars is randomly selected and weighting according the score. Mr. B’s important field selection 200 entries are “mutated” to create a subset of those exemplars by either replacing the index of the associated above-threshold loan or by randomly flipping some number of bits in the bit string. Thereafter, Mr. B’s “mutated” important field selection 200 entries are “crossed-over.” Here, “cross-over” is accomplished by taking a subset of the exemplars by picking pairs of items, and, for each such pair, selecting a single point in the exemplar’s bit string, and exchanging the contents of the bit strings beyond that point. The process is repeated until all possible paths to adequacy are ranked and/or voted according to “contribution” of each field has been computed, and sorted from most influential to least influential.

[0086] It should be noted that the example of Mr. B is a simple and straightforward genetic algorithm, wherein the preferred method has found that the population converges to a set of exemplars that represent changes to fields/variable that produce significant improvements in Mr. B’s creditworthiness (i.e. yielding an acceptable risk profile to issue a loan).

[0087] Referring back to FIG. 2, the fourth and fifth steps are reason text generation 400 and generating an adverse action letter 500.

[0088] There are two aspects of reporting the results of the search. First is reporting the variables that are likely to be wrong (reason text generation 400). Second, reporting the reasons for the adverse action (adverse action letter 500). These are not the same, and solving each requires different mechanisms.

[0089] Reporting the variables that are likely to be wrong (reason text generation 400) is straightforward, given the weighted contributions computed in the previous section. The preferred method for reason text generation 400 involves recording a list of items with the largest possible weights.

[0090] Credit scoring systems often perform veracity checks with third-party data sources that supply information on the borrower. And if a borrower’s profile is inconsistent with what is self-reported and/or has values outside the “norm” of other borrowers, those fields will be flagged, and often result in a deduction from the borrower’s credit score. Thus, there is a strong probability that important errors will show up with high ranks. Since the values associated with those errors and the sources from which the erroneous signals were drawn will be listed, consumers will be able to recognize opportunities for significantly improving their scores by correcting errors in credit agency files or in their own application data.

[0091] At the end of the search and ranking steps, we have one or more “recipes” for transforming a below-threshold loan into an above-threshold one. Unfortunately, the contents of those recipes are not ready to present to anapplicant, as they simply are information of the form “this signal might be associated with a change in the score for your application.”

[0092] Thus, reporting the intelligible reasons for the adverse action letters 500 requires additional steps and procedures. The creation of adverse action letters 500 may be resolved within the standard boundaries of well-studied machine learning paradigms. In essence, the “filtered” field list would then be translated to associated qualitative entries. For example, a variable or meta-variable associated with “number of addresses” would have at least one text entry associated with it (so called “report classes”), such as “your residential address has changed many times in the past five years, indicating that your employment is unstable.”

[0093] Report classes are lender-defined, examples of which include messages that are prescriptive (“Establish and maintain a bank account for more than 2 years” or “Avoid overdrawing your checking account and try to schedule your essential payments so you aren’t late with your bills”), descriptive (“Lexis-Nexis reports have multiple social security numbers associated with your name and address. That could be in error, and, if so, should be corrected.”), and/or monitory (“One or more of the fields in your application exhibits features highly correlated with fraud. You should look at items reported on your application and correct any errors therein.”).

[0094] The preferred method generates a labeled set of training exemplars which connect the weight pattern for a given application to the report class or classes with which the application is associated. Thereafter, the preferred embodiment can use standard classification techniques such as support vector machines, k means, learned vector quantization, or EM to build a labeling function.

[0095] Any of the above-described processes and methods may be implemented by any now or hereafter known computing device. For example, the methods may be implemented in such a device via computer-readable instructions embodied in a computer-readable medium such as a computer memory, computer storage device or carrier signal.

[0096] The preceding described embodiments of the invention are provided as illustrations and descriptions. They are not intended to limit the invention to precise form described. In particular, it is contemplated that functional implementation of invention described herein may be implemented equivalently in hardware, software, firmware, and/or other available functional components or building blocks, and that networks may be wired, wireless, or a combination of wired and wireless. Other variations and embodiments are possible in light of above teachings, and it is thus intended that the scope of invention not be limited by this Detailed Description, but rather by Claims following.

What is claimed is:

1. A central computer server communicatively coupled to a public network, the central computer server having a non-transitory computer-readable medium with a sequence of instructions which, when executed by a processor, causes said
processor to execute an electronic process that automatically generates high quality adverse action notifications, said process comprising:

collecting an electronic dataset for a borrower which contains a credit score and a plurality of variables and meta-variables that describe specific aspects of the borrower to generate a borrower profile;

independently and collectively processing the plurality of variables and meta-variables in the borrower dataset against a lender’s criteria for creditworthiness;

identifying sets of variables and meta-variables in the borrower profile that, when changed, result in an improved measure of creditworthiness, wherein the identifying step includes analyzing at least one shortest path between the borrower dataset and the dataset of at least one of: (i) a perfect applicant and (ii) an average approved applicant; and

generating a report that interprets the at least one shortest path, and variables and meta-variables therein, into plain language through which the borrower may understand how to improve the borrower’s credit score.

2. The central computer server of claim 1, wherein the lender’s criteria for creditworthiness is measured by a credit score.

3. The central computer server of claim 1, wherein the process further includes ranking the identified sets of variables and meta-variables in the borrower profile that, when changed, result in an improved credit score, said ranking using at least one of the following steps:

   * * * *